VOLUME 38 APRIL, 1952 NUMBER 4

PROCEEDINGS

OF THE

NATIONAL ACADEMY OF SCIENCES

OF THE

UNITED STATES OF AMERICA

EDITORIAL BOARD

LINUS PAULING, *Chairman* EDWIN B. WILSON, *Managing Editor*
ALEXANDER WETMORE, *Home Secretary* ROGER ADAMS, *Foreign Secretary*
WILLIAM W. RUBEY, *Chairman of the National Research Council*

J. W. BEAMS
R. E. CLELAND
W. W. RUBEY
HARLOW SHAPLEY
F. E. TERMAN
G. W. BEADLE
E. A. DOISY
G. C. EVANS
C. H. GRAHAM
J. T. PATTERSON
GREGORY BREIT
L. C. DUNN
C. S. MARVEL
J. VON NEUMANN
W. H. TALIAFERRO

Publication Office: Mack Printing Company, Easton, Pa.
Editorial Office and Home Office of the Academy
2101 Constitution Avenue, Washington 25, D. C.

Entered as Second-Class matter, June 21, 1930, at the Post-office at Easton, Pennsylvania, under the Act of August 24, 1912. Accepted for mailing at a special rate of postage provided for in section 1103, act of October 3, 1917. Authorized June 18, 1920.

Published Monthly Price per Annum, $7.50

OFFICERS OF THE ACADEMY

PRESIDENT ..... D. W. BRONK
VICE PRESIDENT ..... EDWIN B. WILSON
FOREIGN SECRETARY ..... ROGER ADAMS
HOME SECRETARY ..... ALEXANDER WETMORE
TREASURER ..... W. J. ROBBINS

INFORMATION TO SUBSCRIBERS

Checks covering subscriptions to the PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES at the rate of $7.50 per annum should be made payable to the National Academy of Sciences, and sent to the Treasurer, National Academy of Sciences, 2101 Constitution Ave., Washington 25, D. C. Single numbers $0.75. These prices effective January, 1951.

Microfilm copies of the PROCEEDINGS are available to subscribers to the paper edition. The microfilm is a positive, furnished on metal reels, suitably labeled and is distributed at the end of the volume year. Inquiries concerning purchase should be directed to University Microfilms, 313 N. First St., Ann Arbor, Michigan.

Past publications of the NATIONAL ACADEMY OF SCIENCES from 1863 to 1926 and of the NATIONAL RESEARCH COUNCIL from 1916 to 1925 are listed in the Index to the First Ten Volumes of these PROCEEDINGS. This Index appeared as Part 2 of Volume 13, No. 1, January, 1927, and is for sale separately at the price of $1.50. Orders should be sent to the Home Secretary, National Academy of Sciences, 2101 Constitution Ave., Washington 25, D. C.

Past volumes of the PROCEEDINGS may be obtained at five dollars per volume unless the sale of the volume would break a complete set of Volumes I to XXX. Single numbers may be obtained for fifty cents except where the sale of such numbers would break up a complete volume. Only forty complete sets of the PROCEEDINGS are available for sale—Volumes I to XXX price $200.00. Orders should be sent to the Home Secretary, National Academy of Sciences, 2101 Constitution Ave., Washington 25, D. C.

THE NATIONAL ACADEMY OF SCIENCES-NATIONAL RESEARCH COUNCIL also issues technical publications reviewing and correlating recent research. The *Bulletin* and *Reprint and Circular Series* have been replaced by a single publication series, available individually, or by subscription at 25% discount. For a list of available publications, with prices and ordering procedures, address the Publications Office, National Academy of Sciences-National Research Council, 2101 Constitution Avenue, Washington 25, D. C.

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES

Volume 38 April 15, 1952 Number 4

MAGELLANIC CLOUDS. III. DIFFERENTIAL ABSORPTION WITHIN THE LARGE CLOUD

By Harlow Shapley and Virginia McKibben Nail

Harvard College Observatory

Communicated February 15, 1952

1. The evidence that much light-dimming material exists in the Large Cloud was presented in an earlier communication[1] where the meagerness of the population of remote galaxies shining through the Cloud was proposed as at least rough proof of the smogginess of that galaxy. At a distance of four or five degrees from the geometrical center of the Cloud, the population of translucent galaxies attains normal richness, even though the outer boundaries are at least six degrees from the center. Outlying parts of the Cloud are, therefore, essentially dust- and gas-free.

The question naturally arises: can we detect the unevenness of the obscuring material in the main body of the Cloud, where the census of faint galaxies is incompetent to reveal different degrees of obscuration? Is it denser around the giant star clusters of the Large Cloud, around the bright nebulosities, and in the Axis, than it is in the open fields?

2. As a by-product of the study of the variable stars, we can now answer these questions. The measuring tool is, of course, the ever useful period-luminosity curve and the Cepheid variables upon which it is based.

With sufficient accuracy for our present purpose, the period-luminosity curve may be written as a straight line

$$\dot{m} = a + b \log P \tag{1}$$

and as

$$\dot{m} = 17.14 - 2.08 \log P \tag{2}$$

for the Cepheids of the Large Cloud that have periods between two and forty days, where P is the period in days and $\dot{m}$ is the median apparent photographic magnitude.[2] Equation (2) applies, therefore, to about 98 per cent of the Cepheid variables. At the currently adopted distance d of 25,000 parsecs, and with an average space absorption over the Cloud of

$\Delta m = 0.4$ mag. (arising partly in our Galaxy and partly in the Large Cloud) we have $m = M + 5 (\log d - 1) + \Delta m = 17.4 + M$, and

$$M = -0.26 - 2.08 \log P \qquad (2')$$

The differential absorption from one part of the Cloud to another appears in the parameter a of equation (1)—that is, it is shown by the vertical shift of the period-luminosity curve as usually plotted. We appropriately assume that there are no significant errors in the determination of period, and that the intrinsic relation of period to luminosity is independent of position in the Cloud.

3. Six regions in the Large Cloud have been selected for thorough investigation. From many intercomparisons of photographs, practically all variables with appreciable variation have been discovered in these fields, except possibly some Algols with relatively narrow minima, and they are more likely to be members of the foregound than of the Cloud. For these

TABLE 1

REGION	LOCATION	AREA IN SQ. DEG.	GALAXIES PER SQ. DEG.	NUMBER OF VARIABLES	SYSTEMATIC DEVIATION	MEAN SQUARE ERROR
A	N 1856	0.67	0	35	+0.27	±0.061
B	Axis	0.75	5	81	+0.09	±0.034
C	30 Doradus	1.0	0	18	+0.18	±0.094
D	Intermediate	2.75	7	28	+0.20	±0.059
E	N 1783	1.0	4	19	−0.13	±0.074
F	N 1866	0.67	9	23	−0.05	±0.039
Totals and Weighted Means			34	204	+0.10	±0.022
A + B	Axis fields		3	116	+0.14	±0.030
E + F	Outer fields		6	42	−0.09	±0.039

six selected regions we are able to report the relative frequency of types of variables, and the relative frequency of the various period lengths of the Cepheids. (See following paper.) Also the period-luminosity curve can be tested in each of these regions for deviation of slope and zero point from the mean. In the following paragraphs the test of zero point is reported.

The six regions that have been studied during the past several years are shown in figure 1 on a reproduction of a photograph of the Large Cloud; they are also listed in the following tabulation. Region A is at the east end of the Axis; Region C, the 30 Doradus neighborhood, has been reported in an earlier communication,[3] and Region F, around NGC 1866, in a paper read before the American Astronomical Society.[4] The new values of periods and median magnitudes will be published in *Harvard Observatory Bulletin 921*.

For all variables in each field of table 1, the deviation of the observed median magnitude from the adopted period-luminosity curve has been derived not graphically but by computing a median magnitude from the

period by means of equation (2). For each field the mean of these deviations, $\dot{m}_o - \dot{m}_c$, and its mean-square error are given in the sixth and seventh columns.

The positive systematic deviations indicate for regions A, B, C and D that the mean median magnitudes of the Cepheids fall below the period-luminosity curve based on the formula. In such places, the light absorption

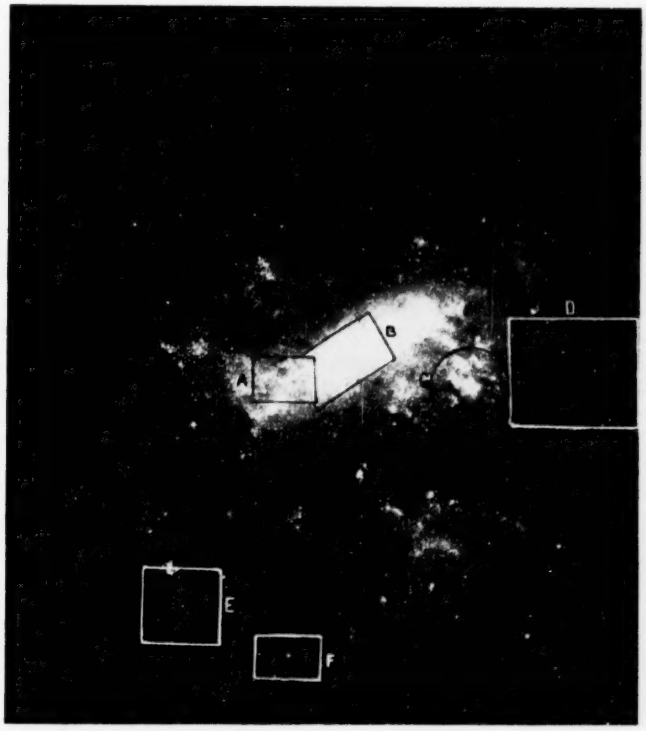

FIGURE 1

Approximate locations and dimensions of the six fields in the Large Cloud.

averages to be heavier than for the Cloud as a whole. Negative systematic deviations indicate greater clarity than the average.

4. Region C contains recognizable bright nebulosities. Presumably dark nebulosity is also present in such an area and doubtless some of the Cepheids studied are imbedded in or are behind the light-dimming nebulosity. On plates made with the 60-inch reflector, as well as on those made with the Bruce refractor, which were used for most of this study, the center and west

end of the Axis is clear of spotted nebulosity, bright or dark, and the small average systematic deviation in field B may reflect this relatively high transparency.

The Intermediate region D is near the great 30 Doradus nebulosity and the large positive deviation suggests that the region may be shadowed by the outlying dark clouds. The two outer regions E and F have negative systematic deviations. They should give the better zero point for the

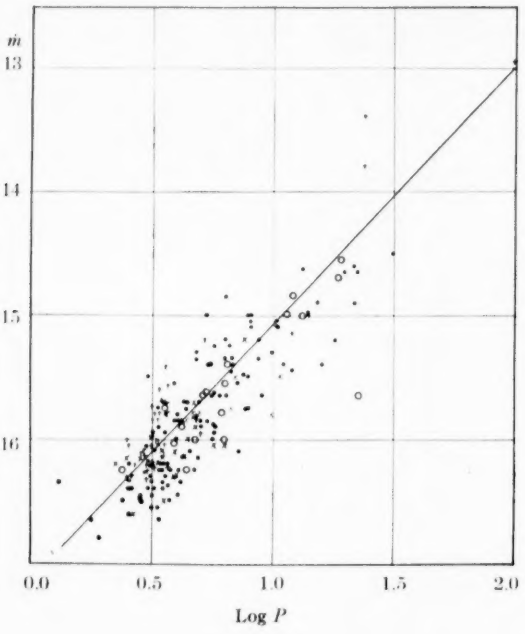

FIGURE 2

The period-luminosity curve. The variables in the outer fields (E, F) are represented as crosses, the intermediate region (D) as X's, the Axis (A, B) as dots, and the 30 Doradus field (C) as open circles.

period-luminosity curve, and, therefore, a preferred value of the distance of the Cloud.

Comparing the two outer regions with the two Axis regions (bottom of table 1), we get a difference $\Delta m = 0.23$, which is a minimum measure of the total absorption. The scarcity of external galaxies in these two outer fields suggests that they also may be strongly affected by obscuring matter, and the total blocking of light in the main body of the Cloud may, therefore,

amount to as much as three-fourths of a magnitude, in addition to that which is contributed from the dust and gas in our galactic system. But at latitude $-33°$ the galactic contribution is, perhaps, negligible, since the frequency of faint galaxies around the Large Cloud is as high as it is at the South Galactic Pole.

5. The material here presented for the six regions provides no sufficient basis for supposing that the period-luminosity curve varies in slope in different parts of the Cloud. Any small difference derived from the data could easily be attributed to the residual errors in the magnitude scale of the secondary sequences used for each field. Especially for this intercomparison of fields, the magnitude sequences have been revised where necessary, and are based on a single master sequence. (Differences in slope between the Large and the Small Cloud are also to be attributed to the differing magnitude scales which we are now adjusting with the aid of photoelectric photometers.)

The formula used above was derived for the Large Cloud some years ago when the periods of only 137 stars were known. They were scattered at random over the Cloud, with a natural selection of those outside the crowded fields. These stars yielded a good period-luminosity curve but not a proper survey of period frequencies. In figure 2 we plot the material from the six thoroughly studied regions (omitting all others), and put in the period-luminosity line of equation 2. There appears to be no need to revise the formula at this time.

6. It should be noted that there are several probable causes of the conspicuous scatter of the individual points. The most important are (1) the spotty light absorption which is shown by the present discussion and is also directly visible on long exposure photographs, and (2) the inevitable dispersion arising from the thickness of the Cloud. For the variables in and beyond the center of the Cloud, the median magnitudes are dimmed both by distance and by absorption; and those on this side of the center are brighter than the average of the same period lengths because of being nearer and free from absorption.

If the diameter of the Large Cloud in the line of sight is comparable with the transverse diameter, then we compute from the approximate formula

$$\Delta m = m_{\text{far}} - m_{\text{near}} = 5 \log [(1 + \sin \theta)/(1 - \sin \theta)]$$

(where θ is the angular semidiameter) that the magnitude difference for Cepheids of a given period on the near and far boundaries (extreme limits) is $\Delta m = 0.48$, aside from the contribution by absorption in the Cloud, and for Cepheids on the near and far edges of the main body of the Cloud, $\Delta m = 0.24$.*

To this very appreciable source of scatter we now add the demonstrated contribution of uneven density of light-dimming material within the Cloud,

as well as the other causes of dispersion in the median magnitudes, such as optical doubling, the Eberhard effect on magnitudes in crowded places, and the errors in measured median magnitudes and calculated periods. Less scattering than we now find is hardly to be expected, even when much more extensive and precise photometry is carried through. Furthermore, it is also quite possible that some Cepheids with identical periods actually have measurably different candlepowers. We have, as yet, no theory that demands exact relationships between period and luminosity, and we have some evidence that the classical Cepheids may not be all of one kind.

* At the distance of the Large Cloud, 1′ is equivalent to approximately 25 light years. The total diameter is 18,000 light years. For the main body of the Cloud the linear diameter is about 10,000 light years, which is probably less than one-quarter of the diameter of the comparable central part of our own galactic system. Consistent with this dimensional difference, the luminosity of the Large Cloud is less than one-tenth of that of the galactic system, and the Cloud's mass and star population may be but a hundredth.

[1] These PROCEEDINGS, **37,** 133–138 (1951); *Harvard Reprint* 345 (1951).

[2] These PROCEEDINGS, **26,** 541–548 (1940); *Harvard Reprint* 207 (1940).

[3] These PROCEEDINGS, **34,** 173–179 (1948); *Harvard Reprint* 306 (1948).

[4] *Ast. J.*, **55,** 249–251 (1951); *Harvard Reprint Series* II-36 (1951).

MAGELLANIC CLOUDS. IV. ON THE PERIOD FREQUENCY ANOMALIES

BY HARLOW SHAPLEY AND VIRGINIA MCKIBBEN NAIL

HARVARD COLLEGE OBSERVATORY

Communicated February 15, 1952

1. One of the most striking facts that has turned up in our studies of the Magellanic Clouds is the preferential concentration of longer period Cepheids into the denser parts of the Small Cloud, while fainter variables, with unusually short periods for classical Cepheids, are dominant in the variable star population of the sparse outer areas. The correlation of average period length with distance from the center of the Small Cloud has been definitely established,[1] but the evidence has not been clear for the Large Cloud. We have now increased the number of its measured variables from 144, published before 1948, to 320, and have made a thorough examination of the period distribution in six regions. (Two-thirds of all the variables in these regions are classical Cepheids.) A new examination of the period frequency anomaly is, therefore, possible. We base it on new data for 204 classical Cepheids, for which the periods of 110 are yet to be published.[2]

Results for the six selected regions are summarized in the following tabulation:

REGION	LOCATION	NUMBER OF VARIABLES	MEDIAN PERIOD	DISTANCE FROM GEOMETRIC CENTER	DISTANCE FROM CENTER OF AXIS
A	N 1856	35	6^d88	$1^\circ5$	$1^\circ2$
B	Axis	81	4.46	1.3	0.3
C	30 Doradus	18	6.24	1.6	1.6
D	Intermediate	28	4.34	2.9	2.9
E	N 1783	19	3.65	3.3	4.3
F	N 1866	23	3.19	2.9	4.2
A + B	Axis fields	116	4.79		
E + F	Outer fields	42	3.26		

The location of the six fields is shown in the photograph of the preceding paper.

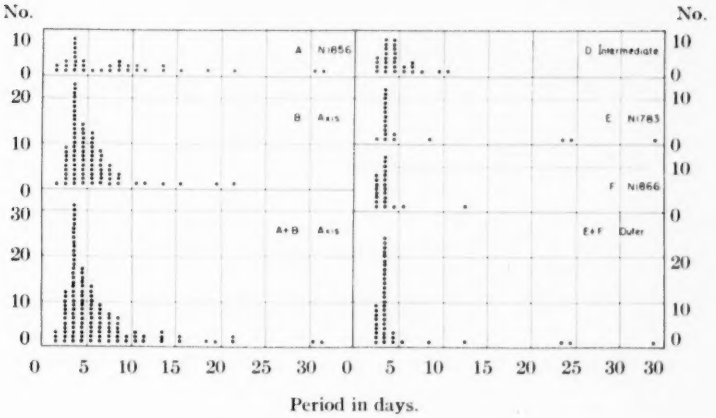

FIGURE 1

Frequency of periods for classical Cepheids in five fields of Large Cloud.

2. The effect clearly shown for the Small Cloud is strongly indicated by the numbers in column 4 of the table. The actual difference between the star-composed "bar" of the Cloud, which we call Axis, and the outer regions is probably greater than shown, since some of the short period variables in our survey of the Axis fields may not be located in the dense regions, but rather are projected on the Axis from the outer parts of the Cloud. Also, we should note that in the most obscured fields (e.g., the east end of the Axis) the faintest variables, which will usually have the shortest periods, would be most easily dimmed out of the reach of our surveys. We believe, however, that this selective effect can by no means erase the observed tendency for the longer periods to appear preferentially in the dense Axis region. The low value of the median for field B makes the anomaly less

convincing here than in the Small Cloud. In field B there are 14 irreguar variables mostly of the type found commonly in the Orion Nebula, and 15 variables of unknown type; most of these are relatively bright and the unrevealed Cepheids among them would move the median toward longer periods.

In figure 1 the distribution of periods is shown for each field separately, and also for the combinations of the Axis pair and the two outer fields.

A preferential nuclear location of long period classical Cepheids in our Galaxy similar to that in the Magellanic Clouds is complicated by the fact that the galactic nucleus is exceedingly rich also in the cluster-type Cepheids which are absent from the Magellanic Clouds.

3. In figure 2 a new comparison is made of the distributions of Cepheid

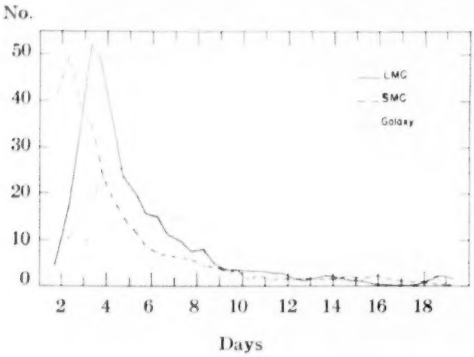

FIGURE 2

The frequency of the periods of classical Cepheids in three galaxies.

periods in both Magellanic Clouds and in the galactic system. The data have been adjusted to equal numbers of stars and have been smoothed in order to bring out pictorially the second important anomaly in the properties of Cepheids in the Clouds—namely, the striking difference in frequency distribution of periods between the galactic system and the Small Cloud, with the Large Cloud intermediate. Only three of the 196* periods of classical Cepheids in the six completely examined fields of the Large Cloud are less than two days. In the Small Cloud 135 periods out of 536 are less than two days. In the galactic system, with data from the Russian Catalogue of 1948, and the 1949 and 1950 Supplements,[3] the score is 20 out of

* The eight stars with periods greater than 20 days are omitted from the figure and computation.

433. The percentages are 1.5, 25.2 and 4.6. The data for the Clouds will change little with increase of numbers, and the ratio of 1 to 17, therefore, stands out as a most remarkable and suggestive phenomenon.

4. That the period frequency anomaly is associated with the stage of development in these similar irregular galaxies is a fair supposition. The Small Cloud, on this hypothesis, is more "advanced" than the Large. It is intermediate in some of its evolutionary characteristics between the dust-and-gas rich Large Cloud and the dust-free Sculptor Cluster, which is a dwarf open-structure spheroidal galaxy very rich in variables of the cluster-type.[4] The Sculptor Cluster has few classical Cepheids and few highly luminous stars; the Small Magellanic Cloud has many of both, but the Large Cloud is much richer in such supergiants.

The scores of Cepheids with periods around 1.5 days in the Small Cloud may be that system's version of cluster-type variation; and this observation leads to the suggestion that the chemistry and the behavior of the Cepheids of a galaxy or globular cluster are related to its mass and stage of development.

[1] These PROCEEDINGS, **28**, 200–204 (1942); *Harvard Reprint* 241 (1941) and *Harv. Obs. Bull.* 916 (1942).

[2] *Harv. Obs. Bull.* 921 (1952), in press.

[3] Kukarkin, B. V., and P. P. Parenago, *The General Catalogue of Variable Stars*, Moscow (1948), *First Supplement to the General Catalogue of Variable Stars*, Moscow (1949), and *Second Supplement to the General Catalogue of Variable Stars*, Moscow (1950).

[4] Shapley, H., *Comparison of the Magellanic Clouds with the Galactic System*, Mich. Pub., **10**, 79–84 (1951); *Harvard Reprint Series* II–37 (1951).

*BREAKAGE OF CHROMOSOMES BY OXYGEN**

BY ALAN D. CONGER AND LUCILE M. FAIRCHILD

BIOLOGY DIVISION, OAK RIDGE NATIONAL LABORATORY OAK RIDGE, TENNESSEE

Communicated by Karl Sax, February 12, 1952

The general physiological effects of oxygen on tissues and whole organisms, such as non-nuclear effects, death and other phenomena, are well known. Cleveland[1] has mentioned that oxygen will partially or completely destroy the chromosomes and nuclear membrane of the gametophyte of the flagellate *Trichonympha*, which lives in a roach. This paper, however, is concerned only with a very specific effect of oxygen on the cell, the breakage of the chromosomes and production of chromosomal aberrations identical with those induced by radiation.

Treatment of Tradescantia flower buds and dry pollen grains with oxygen produces chromosomal aberrations which can be observed in the

pollen grain or pollen tube division. Chromosomal aberrations produced by oxygen are in all respects indistinguishable from those found after exposure to ionizing radiations. The magnitude of the effect is considerable, from two to ten times as great as is generally studied in radiation experiments. Methods of treatment, culture and subsequent observation of broken chromosomes are identical with those employed in conventional radiation experiments, except that oxygen or some other gas is used instead of radiation.

Some of the features of the breakage of chromosomes by oxygen will be presented in the following sections. Before presenting the detailed material we may say that, to date, we have produced aberrations in pollen grains no less than nineteen different times, in about as many ways, and over a period of four months. Also, though there may be some quantitative variation between experiments from week to week, most experiments are consistent among themselves.

The experiments have revealed that the chromosome breakage is due to the oxygen itself, not the impurities in it; that light is not necessary in the reaction; that the magnitude of effect is related to time of exposure, oxygen concentration and the pressure at which it is applied.

Methods.—Gas has been applied to pollen or other material in a variety of ways by: (*a*) flushing gas over the pollen at normal pressure, (*b*) evacuating with a Megavac pump (to a pressure of about 100 μ of Hg) and admitting gas to normal pressure a number of times, (*c*) admitting gas to a small pressure bomb to positive pressures and releasing to normal pressure a number of times or (*d*) combinations of these treatments.

The bulk of the experiments to be described were made on dry pollen grains of the flowering plant *Tradescantia paludosa* Anderson and Woodson (Sax clone No. 5). These dry, air-borne ellipsoidal cells are about 40×20 μ and have a moderately thickened cell wall. The pollen is collected from freshly opened flowers in the morning, desiccated for an hour or so over calcium sulphate to facilitate handling, then dusted onto a glass slide in a layer only a few cells thick. The very thin layer of pollen is exposed to gas in either a small brass pressure bomb or in glass tubes with stopcocks at either end; gas is flushed through the containers from end to end, across the pollen. The treated pollen grains are sowed on a lactose-agar-colchicine medium spread on a microscope slide and cultured in a moist box for about 18 hours (modified from the method of Bishop, 1949[2]). Colchicine is added to shorten chromosomes and accumulate nuclei at metaphase of division. The slides are fixed, stained by the Feulgen technique, and made into permanent mounts. Chromosomal aberrations are observed in the generative nucleus at the pollen tube division. Flower heads of Tradescantia were handled similarly, and entire growing plants were exposed to gas under bell jars. Commercial cylinders of gas from various suppliers (Linde

oxygen >99.5% pure and nitrogen >99.7% pure, Matheson medical oxygen >99.8% pure and Southern Oxygen Co. helium >99.8% pure) have all been employed. One of the tanks of oxygen used for most of the experiments was analyzed at this laboratory and found to be >99.5% pure.

Experiments and Results.—In the following experiments, the breakage of chromosomes by oxygen under different conditions will be described. For each of the experiments, an appropriate control was made, identical to the oxygen treatment except that no oxygen was used. The control data showed that no chromosomal effect was caused by the methods of pollen culture, the conventional handling of the pollen or by repeated evacuations or pressure applications to the pollen. Since all the control experiments were negative, the individual data are presented as a total only in table 1.

TABLE 1

SPONTANEOUS ABERRATION FREQUENCY IN POLLEN

Total of one or more control treatments from each experiment

Experiments	8
Slides	32
Total Cells	1089
Normal Cells:	
No.	1060
Fraction	0.972
Aberrations:	
Chromatid plus isochromatid deletions	
No.	29
Per cell	0.027
Exchanges	
No.	1
Per cell	0.001

One of the early experiments indicated that aberrations were produced in about the same amount whether oxygen was applied to the cells by merely flushing at normal pressure or by repeated evacuation and gas admission, at least when the period of treatment was not less than $1^1/_2$ hours. The experiment also shows that the effect is caused by oxygen from two sources. These data are given in table 2, part A.

That this effect is being caused by the 0.4% or less of impurities present in oxygen from commercial cylinders seems unlikely since the usual contaminants—moisture, hydrogen, and nitrogen—are biologically inert. Nevertheless, a test was made using a 1-liter flask of reagent grade oxygen (100% pure, no contaminants) sealed to an acid-cleaned, degassed, all-glass system. After inserting the pollen into the system and evacuating for three minutes, the system was closed off from the pump and the glass seal to the oxygen flask broken. Pollen remained in the oxygen for 2 hours. Data are in table 2, part B.

TABLE 2

ABERRATION PRODUCTION BY OXYGEN UNDER VARIOUS CONDITIONS

DATE	GAS	PER CENT	PRESSURE[a]	TIME IN GAS (HRS.)	TREATMENT	NO. SLIDES	NO. CELLS	NORMAL CELLS NO.	NORMAL CELLS FRACT.	ABERRATIONS CD. NO.	ABERRATIONS CD. PER CELL	ABERRATIONS ISOCD. NO.	ABERRATIONS ISOCD. PER CELL	EXCHANGES PER CELL
					A. Flush and Evacuation									
June 7	O_2	100	N	1½	Vac. → gas, 7×, tank I	4	65	1	0.02	153	2.4	99	1.5	
	O_2	100	N	1½	Vac. → gas, 7×, tank II	3	50	2	0.04	134	2.7	70	1.4	
	O_2	100	N	1½	Flush only, 1½ hrs.	4	55	2	0.04	145	2.6	88	1.6	
					B. Reagent Grade Oxygen									
Aug. 14	O_2	100	N	2	Vac. → gas, 1×; 100% pure O_2	7	100	6	0.06	> 7 fragments/cell, average				
					C. Positive Pressures									
July 5	He	100	20 lbs.	1½	20 lbs. → normal pressure 10×	5	188	185	0.98	3	0.02	..	..	
	O_2	100	20 lbs.	1½	20 lbs. → normal pressure 10×	2	50	6	0.12	60	1.2	61	1.2	
	O_2	100	96 lbs.	1½	96 lbs. → normal pressure 10×	3	29	0	0	> 20 fragments/cell, average				

D. Treated in Dark

Aug. 23	O_2	100	N	1½	Cells in dark ½ hr. before during and 3 hrs. after gas	3	60	1	0.02	~5 fragments/cell, average			

E. Compressed Air, 5 Atm.

Aug. 9	C. A.	100	60 lbs.	¾	60 lbs. → normal pressure 10×	4	200	119	0.60	95	0.48	45	0.22
Aug. 23	C. A.	100	60 lbs.	1½	60 lbs. → normal pressure 10×	3	100	42	0.42	84	0.84	40	0 .40
Aug. 27	C. A.	100	60 lbs.	1½	60 lbs. → normal pressure 10×	6	60	40	0.67	32	0.53	12	0.20
	O_2	100	N	1½	Vac. → gas, 10×	5	60	9	0.15	79	1.32	33	0.55

F. Time of Exposure

Aug. 27	..	..	..	....	Control total, table 1	32	1089	1060	0.97	29	0.03	1	0.001
	O_2	100	N	10 sec.	Vac. → gas 10 sec. 1×	6	200	192	0.96	12	0.06	..	..
	O_2	100	N	50 sec.	Vac. → gas 10 sec. 5×	6	200	177	0.88	32	0.16	6	0.03
	O_2	100	N	100 sec.	Vac. → gas 10 sec. 10×	6	150	140	0.82	18	0.12	3	0.02
	O_2	100	N	1½ hrs.	Vac. → gas 10 sec. 10×	5	60	9	0.15	>6 fragments/cell, average			

[a] N = normal atmospheric pressure. 20 lbs. = 20 lbs. gauge pressure = 35 lbs. absolute.

This experiment would seem to establish the fact that it is the oxygen and not the 0.4% of contaminants present in commercial cylinders that is causing this effect. In experiments like this one with long exposure times or with high pressures, the chromosomes are so drastically affected that only fragments are seen in most dividing nuclei, as many as 50 in some. The nuclei resemble those found in cells irradiated with 1000 r or more of x-rays. At these very high levels of effect it is difficult to score individual aberrations accurately and the frequency of "fragments per cell" is a more reliable measurement. This high fragment frequency is observed again at the highest pressure in a later experiment using positive pressures of two to seven atmospheres delivered to the pollen on a glass slide inside a brass pressure bomb. Data are in table 2, part C.

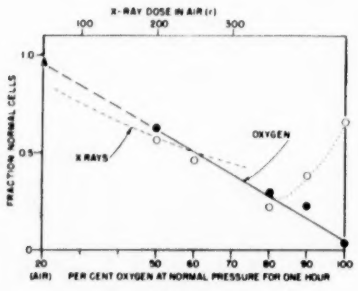

FIGURE 1

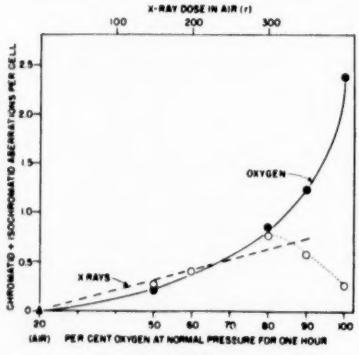

FIGURE 2

Figure 1. Fraction of normal (aberration free) cells against oxygen concentration, and against X-ray dose in air.

Figure 2. Frequency of chromatid plus isochromatid aberrations against oxygen concentration.

Symbols: ● 21 June; ○ 25 June; ▲ Control (air).

It is seen from this experiment that as the oxygen pressure is increased the effect also increases. That the pressure itself is not causing this is demonstrated by the lack of effect when helium is used.

Since these pollen grains are relatively small transparent cells and are known to transmit light, there was a reasonable probability that the breakage of chromosomes observed was due to some sort of combined effect of oxygen plus light. The possibility that this is a photoactivated-oxygen reaction was tested by an experiment in which the pollen was kept in complete darkness for the period from $^1/_2$ hour before the commencement of the oxygen treatment to three hours after its completion and the removal of the cells from oxygen. A brief period, less than five minutes, of illumi-

TABLE 3

RELATION OF ABERRATION PRODUCTION TO OXYGEN CONCENTRATION, AT NORMAL PRESSURE. MIXTURES OF OXYGEN/NITROGEN

DATE	TREATMENT	OXYGEN, %	NO. SLIDES	NO. CELLS	NORMAL		ABERRATIONS			
					NO.	FRACT.	CD. + ISOCD. NO.	CD. + ISOCD. PER CELL	EXCHANGES NO.	EXCHANGES PER CELL
June 21	Vacuum → normal pressure 5×; flush 5 min. in gas 1 hr.	100	5	46	2	0.04	101	2.4	58	1.2
		90	2	50	12	0.24	62	1.2	38	0.76
		80	3	44	13	0.30	38	0.86	17	0.39
		50	5	96	60	0.62	21	0.22	14	0.14
June 25		100	1	50	33	0.66	13	0.26	12	0.24
		90	2	59	23	0.39	34	0.57	25	0.42
		80	2	58	13	0.22	44	0.76	41	0.71
		60	2	68	31	0.46	22	0.40	29	0.43
		50	2	100	56	0.56	24	0.24	34	0.34
		X-RAY DOSE, R								
May 8	X-rays 250 kvp., 330 r/min. in air N.T.P.	50	3	200	177	0.88	16	0.08	5	0.02
		100	4	240	178	0.74	57	0.24	12	0.05
		300	6	315	143	0.45	184	0.58	52	0.16

nation (by a red darkroom safelight only) was necessary for sowing the pollen on the slides. After this, the cells were grown in total darkness until the slides were fixed and stained 18 hours later. Results are given in table 2, part D, which show that the oxygen effect is not dependent upon a photoactivated process.

The relation of aberration production to oxygen concentration, at a constant time, has been tested twice. Oxygen/nitrogen gas mixtures after repeated evacuations were flushed over the pollen for one hour. Data are in table 3, which includes for comparison results of X-irradiation in air. The same data are plotted in figures 1 and 2.

The graphs illustrate several features of the oxygen effect. The two points at 90 and 100% oxygen on June 25 (indicated by the dotted line departing from the oxygen curve) show the occasional quantitative variation found between supposedly identical experiments done at different times In this particular case, we suspect that the mixtures for the two deviating points were improperly prepared or some back-flow of gases between tanks occurred during the mixing process. If these two points be ignored, the consistent data of June 21 combined with the remainder of the June 25 data seem to indicate rather clearly several things: (1) As oxygen concentration (at normal pressure) is increased, the effect increases. (2) The curve of normal cells against concentration, if extended back, falls to approximately zero effect at 20% oxygen (air). (3) Chromatid plus isochromatid deletions appear to be increasing as a power of oxygen concentration. (4) An equal effect is produced by 65% oxygen for one hour and 350 r of x-rays.

The last experiment suggests that any partial pressure of oxygen in excess of that normally found in air, whether achieved by concentration or by pressure, should cause chromosome breakage. This has been verified by using compressed air at positive pressures, table 2, part E. Though it has been tested but once, on August 27, it appears that air (20% oxygen) at 5 atmospheres does not produce as much affect as 100% oxygen at one atmosphere. In addition, these results with compressed air confirm the evidence with reagent grade oxygen that it is not the contaminants present in oxygen cylinders that are causing the effect.

The previous experiments indicate that effect is a function of oxygen concentration and pressure, and we may assume of time also. A preliminary test of the relation of effect to time of exposure (concentration and pressure constant), table 2 part F, illustrates this. Though the data are small, it appears that normal cells are disappearing exponentially with time. It is interesting that the effect is still increasing up to an hour or more. Theoretically, the relaxation time for oxygen diffusion to the center of these cells (considered as a sphere of water) is less than 0.1 second. The very great difference between these times indicates that for oxygen

TABLE 4

ABERRATION PRODUCTION BY OXYGEN IN THE FLOWER BUDS OF *Tradescantia*

Experiments:

Sept. 6—Control, inflorescences taken from greenhouse plants.

Aug. 24, Aug. 31—Plants grown in O_2 for 4 days. Slides made when removed from oxygen.

Sept. 5—Inflorescences in O_2 at 100 lbs. for 3 hrs.; slides made 24 hrs. later.

EXPT.	TREATMENT	OBSERVED	NO. SLIDES	NO. CELLS	NORMAL NO.	NORMAL FRACT.	ABNORMAL NO.	ABNORMAL PER CELL
		A. Observed at Meiosis in Pollen Mother Cells						
Sept. 6	Control	Chromo. fragments at ana. I	2	200	191	0.97	9	0.045
Aug. 31	O_2 4 days	Fragments and bridges at ana. I and II	3	150	111	0.74	39	0.26

EXPT.	TREATMENT	OBSERVED	NO. SLIDES	NO. CELLS	NORMAL NO.	NORMAL FRACT.	CD. + ISOCD. NO.	CD. + ISOCD. PER CELL	EXCHANGES NO.	EXCHANGES PER CELL
		B. Observed at Mitosis in Microspores								
Sept. 6	Control		5	250	250	1.00	0	0	0	0
Aug. 24	O_2	Plant 1	1	50	45	0.90	5	0.10	0	0
	4 days	Plant 2*	4	212	169	0.80	65	0.31	20	0.10
		(One bud of Plant 2)	1	12	2	0.17	30	2.50	11	0.92
Sept. 5	O_2, 3 hrs. at 100 lbs.		1	80	76	0.95	3	0.04	1	0.01

* Three chromosome dicentrics were also found.

to exert its effect it must do more than simply diffuse into the cell and reach the nucleus.

A number of attempts have been made to produce chromosomal aberrations with oxygen in material other than the dry pollen grains. It has been possible to do this to the microspores and pollen mother cells in the flower buds of Tradescantia by growing entire plants for four days in an atmosphere of oxygen with a slow constant flow of gas, and by a 3-hour exposure of inflorescences to oxygen at 100 pounds pressure. Results are given in table 4 which show that oxygen can break chromosomes in normal, wet, metabolizing tissues in both prophase and resting stage, at meiotic and mitotic divisions, and with treatment periods as short as three hours.

We would like to emphasize here that, to date, our detection of oxygen-broken chromosomes in the flower buds has been exceedingly variable. We have generally observed in these flower bud experiments that the majority of the treated buds die before the cells can be examined and are therefore unavailable for chromosomal observation, and the surviving buds are outwardly unaffected and show no broken chromosomes. In the few buds, observed before death of the cells is complete, multitudes of chromosomal aberrations identical to radiation induced ones are observed. From our experience, we are reasonably convinced that oxygen-broken chromosomes are not observed more frequently because in wet tissues there is a very close tolerance between breaking chromosomes and killing the cell, and usually the latter effect is produced. It is rather striking that a treatment as short as three hours will produce this effect.

Discussion.—Several of the features of chromosome breakage by oxygen indicate that the cell has an upper limit of oxygen tolerance which if exceeded, by an amount that is a function of the oxygen partial pressure and time of exposure to it, will cause chromosomal breakage. That this is a metabolic function of the cell seems clear from the fact that no effect is caused by air (20% oxygen) and that there is an increase in effect with time of exposure up to an hour, much longer than the simple diffusion time. The cell apparently has some normal rate for oxygen metabolism, and the amount of chromosomal breakage induced is related to the amount by which this oxygen rate is exceeded.

Although it is stated that the effect falls to zero at a partial pressure of 20% oxygen (air), the control results in table 1 show that about 3% of the pollen grains normally have "spontaneous" chromosome aberrations. The experiment on oxygen concentrations suggests that this residual 3% of spontaneous breaks may be due to the oxygen normally present in air; the conclusion is particularly suggestive since it has been shown that the natural radiation level is about a thousand-fold too small to account for the spontaneous chromosome breakage rate.[3]

The identity of the effect of oxygen and ionizing radiations on these chromosomes, oxygen producing all of the same type of aberrations as radiations, implies that the two agents are acting in a similar way. Radiation is known to produce active radicals in water, of which OH and HO_2 are the most suspect in biological effects.[4] Oxygen may be producing its effects on the chromosomes through the action of these same radicals, some of which can be formed by certain oxidative enzyme systems known to occur in plants. With the present limited data, it would be fruitless to speculate further on this point.

It seems fair to say that the breakage of chromosomes by oxygen is probably a phenomenon of general application to normal tissues, and is not restricted to specialized cells like the dry pollen grains, although it is more readily produced in them. The positive results in causing breakage in the flower buds, in several stages of division, demonstrate this. But from our numerous attempts to cause chromosomal breakage in the flower buds with oxygen, mostly resulting in the death of the cells, we suspect that in wet tissues an oxygen treatment sufficient to break chromosomes is generally lethal (or almost so) to the cell.

Summary.—Chromosomes have been broken in the dry pollen grains and in the microspores of the plant Tradescantia by exposure of the cells to partial pressures of oxygen greater than that of air. The chromosome aberrations produced are identical to those caused by ionizing radiations. Exposure to 100% oxygen for one hour will produce as many chromosomal aberrations as approximately 1200 r of x-rays; 65% oxygen, as much effect as about 350 r.

The aberrations are not being caused by the methods of treatment employed. The reaction is not photoactivated. Magnitude of effect is related to time of exposure and to partial pressure of oxygen, increasing from almost no effect in air (20% oxygen) to maximum measurable effect at 100% oxygen. Increasing the partial pressure of oxygen by exposing cells in ordinary air at positive pressure will break chromosomes also.

* Work performed under Contract No. W-7405-eng-26 for the Atomic Energy Commission.

[1] Cleveland, L. R., "The Whole Life Cycle of the Chromosomes and Their Coiling Systems," *Trans. Am. Phil. Soc.*, new series **39,** 1–100 (1949).

[2] Bishop, C. J., "Pollen Tube Culture on a Lactose Medium," *Stain Technol.*, **24,** 9–12 (1949).

[3] Giles, N. H., "Spontaneous Chromosome Aberrations in Tradescantia," *Genetics*, **25,** 69–87 (1940).

[4] Radiation Chemistry Symposium, London, 1950, *Brit. J. Radiology*, **XXIV,** 413–440 (1951).

MUTANT ISOALLELES AT THE VERMILION LOCUS IN DROSOPHILA MELANOGASTER

By M. M. Green

Division of Genetics, University of California, Davis, California

Communicated by R. E. Clausen, February 6, 1952

The term isoalleles has been used to describe allelic genes which when homozygous product phenotypes indistinguishable from one another, but which can be separated from each other by special tests.[1] Wild type isoalleles have been described for the cubitus interruptus[1] and white[2, 3] loci in *Drosophila melanogaster* and for the bobbed locus in *D. hydei*.[4] Similar cases which may be ascribed to isoallelism are the multiple six alleles in Lymantria and Habrobracon and the self-sterility alleles in Nicotiana, Oenothera, etc. The results to be reported herein, viz. isoallelism among vermilion mutants in *D. melanogaster* are of interest because of the known biochemical basis for the action of vermilion mutants. Thus it should be possible to attempt a more exact analysis of the nature of isoallelism.

Mutants at the vermilion locus in *D. melanogaster* are recessive, sex-linked and are characterized phenotypically by the almost complete absence of the brown pigment component of the eyes. From the analysis of the behavior of vermilion mutants in gynandromorphs,[5] of transplantation of optic disks[6] and later of biochemical studies[7, 8] it has been shown that in vermilion mutants a failure occurs in the conversion of tryptophane to kynurenine. This inability to convert tryptophane to kynurenine in vermilion flies results in a block in the synthesis of brown eye pigment. When vermilion eye disks are supplied kynurenine—as in gynandromorphs, in transplants into wild-type hosts, or by feeding or injection of kynurenine to vermilion larvae—brown pigment synthesis occurs. In addition it has been found that brown pigment formation occurs in homozygous vermilion flies made homozygous for certain suppressor mutants.[9] As will be demonstrated here, by using the suppressor mutants it is possible to distinguish between vermilion mutants of independent origin which are otherwise phenotypically inseparable.

The following vermilion mutants were used (mutants which arose spontaneously will be designated by (s), and mutants which arose as a result of x-ray treatment will be designated (X)): v^1(s), v^2(s), v^{36f}(s), v^{48a}(X), v^{51a}(X), v^{51b}(s), v^{51c}(X). The various *v* mutants were tested *inter se*, and it was observed that compound ♀ ♀ carrying any two of the *v* mutants were clearly vermilion in phenotype. Two sex-linked, recessive suppressor mutants were used: su^2-*s*, a suppressor of sable and of vermilion and spontaneous in origin, and su^v-*v*, an x-ray induced suppressor of vermilion. The usual tests showed su^2-*s* and su^v-*v* to be allelic.

To determine the phenotype of the eyes of the various *v* mutants in the presence of the suppressors the following procedure was used. The suppressors are localized at position 0.0 on the X chromosome and are therefore completely linked with the recessive yellow (*y*) body color. The *v* mutants are localized at 33 on the X chromosome and thus very closely linked to raspberry (*ras*), another recessive eye color mutant located at 32.8. The eye colors of *v*, *ras* and *ras v* are all easily separable phenotypically. Stocks of the genotypes su^2-*s*, ras^2v^1 and su^g-*v*, ras^2v^1 were constituted. These carry the y^+ allele of *y*. Since v^1 is suppressed by both suppressors, the aforementioned flies are *ras* v^+ and y^+ in phenotype, and thus *ras* serves to mark the presence of v^1, and y^+ the presence of the suppressor. Stocks of the various *v* mutants marked with the mutant *y* and lacking either suppressor were also made up. By the appropriate crosses, ♀ ♀ of the genotype su^2-*s*, $ras^2v^1/y\ v$ were obtained and these were crossed to sibling ♂♂. The male progeny were checked. Single crossovers between su^2-*s* and ras^2

TABLE 1

PHENOTYPES OF *v* ALLELES IN TESTS WITH SUPPRESSOR MUTANTS AND IN GYNANDROMORPHS

	EYE PHENOTYPE		
	HOMOZYGOUS	*v* WITH	
v ALLELE	HOMOZYGOUS su^2-*s*	HOMOZYGOUS su^g-*v*	IN GYNANDORMORPHS
v^1	Wild-type	Wild-type	Non-autonomous
v^2	Wild-type	Wild-type	. . . . .
v^{36f}	*v*	*v*	Non-autonomous
v^{48a}	*v*	*v*	. . . . .
v^{b1a}	*v*	*v*	Non-autonomous
v^{b1b}	*v*	*v*	Non-autonomous
v^{b1c}	*v*	*v*	Non-autonomous

result in ♂♂ of the genotype *y* ras^2v^1 and su^2-*s*, *v*. Since crossovers between ras^2 and v^1 are relatively rare, and may be disregarded, all ♂♂ phenotypically y^+ras^+ must be genotypically su^2-*s*, *v*, and examination of these ♂♂ permits the determination of whether the suppressor suppresses the *v* mutant being tested. Since crossing over between su^2-*s* and ras^2 is high, ca. 30%, these ♂♂ are abundant. Following recovery of su^2-*s*, *v* ♂♂, stocks were constituted in which ♀ ♀ homozygous su^2-*s*, *v* were obtained. Identical tests were made using su^g-*v*, ras^2v^1.

The results of these tests are listed in table 1. It will be noted that the *v* mutants fall into two classes: those that are suppressed (collectively designated as v^s) in which the *v* phenotype is modified to wild-type in the presence of either suppressor mutant, and those that are unsuppressed (collectively designated as v^u) in which the *v* phenotype is unaltered in the presence of either suppressor mutant. The phenotypic separation of the *v* mutants in the presence of the suppressors is unequivocal. Phenotypes

of the ♀ ♀ agreed without exception with that of the ♂ ♂ for each *v* mutant tested. Among the v^u mutants are those of both spontaneous and x-ray origin. These results allow certain assumptions to be made regarding the nature of the action of *v* mutants. It may be assumed that in the v^s mutants, the presence of the suppressor allows the conversion of tryptophane to kynurenine to take place. Once this metabolic block is released, brown pigment synthesis proceeds. Conversely, it may be assumed that in the case of the v^u mutants the suppressor is ineffective in promoting the conversion of tryptophane to kynurenine. There is evidence from Neurospora that suppressor mutants release specific biochemical blocks.[10]

Proceeding from these assumptions, it is necessary to provide an explanation for the failure of the suppressor to act in the presence of the v^u mutants. One hypothesis which suggests itself is that the v^u mutants lead to a complete inactivation of the brown pigment synthesis system. In such a scheme the suppressors might foster the conversion of tryptophane to kynurenine but v^u individuals cannot utilize the kynurenine or any other component of the brown eye pigment mechanism. Since it is known that v^1 eyes develop non-autonomously in gynandromorphs[5] the hypothesis can be tested by observing the behavior of v^u mutants in gynandromorphs. If the total inactivation mechanism operates then it would be predicted that v^u mutants would be autonomous. Gynandromorphs involving v^u mutants were obtained from the following cross: ♀ $y\ sn^3\ v^u$ X ♂ X^{c2} (ring X chromosome carrying $y^+sn^+v^+$). The use of the X^{c2} chromosome apparently increases the occurrence of gynandromorphs.[11] F_1 progeny were examined and numerous gynandromorphs were obtained. Study of the gynandromorphs showed that the genotypically *v* eye was modified phenotypically to wild type. Thus the v^u mutants behave non-autonomously. These observations eliminate the possibility of a total inactivation mechanism being involved.

An alternative hypothesis stems from more recent information on the mechanism of the conversion of tryptophane to kynurenine. By a most elegant series of experiments, it was demonstrated that in rat liver the step-wise conversion of tryptophane to kynurenine involves two intermediate compounds of which only the second, formylkynurenine, has been identified.[12, 13] These findings suggest that perhaps the difference between v^s and v^u mutants rests in their effect on different steps in the conversion of tryptophane to kynurenine. Thus v^s mutants might block the conversion of tryptophane to formylkynurenine and v^u mutants might block the conversion of formylkynurenine to kynurenine. Accordingly the suppressors, which might be assumed to act in releasing only a specific biochemical block, would foster the conversion of tryptophane to formylkynurenine and thereby suppress v^s mutants but could not suppress v^u mutants which are blocked biochemically at a different step.

To test this hypothesis experiments were conducted in which formylkynurenine and kynurenine were fed to v^u and v^s mutants. Three v mutants, v^1, v^{36f} and v^{51c} were used. All three mutants were combined with the recessive, autosomal eye color mutant brown (*bw*) such that all *v; bw* homozygotes had eyes essentially white in appearance. Ten larvae of each genotype of the age 72–96 hrs. after oviposition were added to 2 ml. of regular corn meal-molasses-agar medium supplemented with either 1 mg. *l*-kynurenine sulfate or 1 mg. formyl-*dl*-kynurenine. Larvae grown on unsupplemented medium served as controls. In addition, control and test media were supplemented with 50 μg. of streptomycin in order to eliminate the possibility of bacterial conversion of tryptophane to kynurenine. Tests were run in duplicate at 22–24°C. The results are tabulated in table 2. It is evident that both formylkynurenine and kynurenine act as precursors for brown pigment in Drosophila. No difference was noted in the level of brown pigment formed by flies fed either compound, nor was any

TABLE 2

EFFECT OF FEEDING KYNURENINE AND FORMYLKYNURENINE ON THE EYE COLOR OF *v* MUTANTS, *D. melanogaster* AND *v* AND *cd* MUTANTS, *D. virilis* (SEE TEXT FOR DETAILS OF METHOD)

GENOTYPE	PHENOTYPE		
	CONTROL	KYNURENINE	FORMYLKYNURENINE
v^1;*bw*	0[a]	++	++
v^{36f};*bw*	0	++	++
v^{51c};*bw*	0	++	++
v^{40d};*es*	0	+	+
cd;*es*	0	+	+

[a] Estimate of amount of brown eye pigment formed where 0 represents no pigment and ++++ represents the apparent pigmentation of homozygous *bw*.

difference noted among the three *v* mutants tested. If the v^s and v^u mutants act by blocking tryptophane metabolism, then the blocks must come prior to the formation of formylkynurenine. It has been shown previously that the v^1 mutant accumulates tryptophane.[14] Since it has been demonstrated that there is an unknown intermediate compound formed between tryptophane and formylkynurenine,[12, 13] it might be postulated that v^s mutants block tryptophane metabolism between tryptophane and the intermediate and v^u mutants block between the intermediate compound and formylkynurenine. It would then follow that the suppressor mutants release the block between tryptophane and the intermediate compound. The precise chemical nature of the intermediate compound is not clear, although there is strong evidence that it is not alpha oxytryptophane.[15] Consequently the obvious tests cannot be performed.

An additional hypothesis to explain the action of the suppressor might postulate that the suppressor directs the metabolism of tryptophane along

a route not involving the formation of formylkynurenine. Such a hypothesis would presume further that the alternate pathway occurs only in v^s mutants. Evidence to date has not elucidated an alternative pathway,[15] and such a hypothesis appears unlikely.

The observation that formylkynurenine acts as a precursor for both v^s and v^u suggested that similar tests be made on two mutants in *D. virilis*, vermilion (*v*), sex-linked and recessive and cardinal (*cd*), autosomal and recessive. It has been reported that both mutants are non-autonomous when transplanted to wild-type hosts and when larvae of both are fed kynurenine.[16] Tests were performed in the manner noted above using the *v* allele, v^{40d} and *cd*. Both v^{40d} and *cd* were first made homozygous for the recessive eye color mutant eosinoid (*es*), the homologue of *bw* of *D. melanogaster*. The results as listed in table 2 demonstrate that formylkynurenine acts as a precursor to brown pigment in *D. virilis* and that the block in pigment synthesis in both mutants occurs prior to the formation of formylkynurenine. No difference was observed in the level of brown pigment formed by flies fed either compound nor was any difference observed between the two mutants studied.

The results reported demonstrate that in *D. virilis* formylkynurenine functions as a precursor for two separate mutants, v^{40d} and *cd*, while in *D. melanogaster* for only *v* mutants. There is evidence strongly indicating homology between the *v* mutants of the two species[17] but thus far no homolog of *cd* has been uncovered in *D. melanogaster*. It is tempting to postulate that the *v* mutants in *D. melanogaster* are not truly allelic, but rather the v^s and v^u mutants represent a case of close linkage (pseudoallelism) much like that reported for the mutants Star and asteroid[18] and the lozenge mutants[19] in *D. melanogaster*. It would follow from such a postulate that one class of *v* mutants (perhaps v^u) of *D. melanogaster* is homologous to *v* of *D. virilis*, while the other (perhaps v^s) is homologous to *cd* of *D. virilis*. This explanation can remain for the present only as a postulate. The validity of a linkage hypothesis can be established only by demonstrating crossing over between v^s and v^u mutants. Experiments to test the possibility of close linkage are now in progress.

Summary.—(1) The use of suppressor mutants makes possible the classification of *v* mutants in *D. melanogaster* into two types: suppressed and unsuppressed. (2) Formylkynurenine is utilized as a precursor for brown eye pigment in Drosophila. (3) The mutants investigated, *v* in *D. melanogaster* and *v* and *cd* in *D. virilis* appear to be blocked biochemically between tryptophane and formylkynurenine.

Acknowledgment.—The author desires to express his thanks to Dr. C. Yanofsky and Dr. S. Lepkovsky for the kynurenine and to Dr. C. E. Dalgliesh for the formylkynurenine used in the experiments reported here.

[1] Stern, C., and Schaeffer, E. W., PROC. NATL. ACAD. SCI., **29**, 361–367 (1943)
[2] Timofeeff-Ressovsky, N. W., *Biol. Zbl.*, **52**, 468–476 (1932).
[3] Muller, H. J., *J. Genet.*, **30**, 407–414 (1935).
[4] Spencer, W. P., *Genetics*, **29**, 520–536 (1943).
[5] Sturtevant, A. H., *Proc. Sixth Intern. Cong. Genetics*, **1**, 304–307 (1932).
[6] Beadle, G. W., and Ephrussi, B., *Genetics*, **21**, 225–247 (1936).
[7] Butenandt, A., Weidel, W., and Becker, E., *Naturwiss.*, **28**, 63–64 (1940).
[8] Tatum, E. L., and Haagen-Smit, A. J., *J. Biol. Chem.*, **140**, 575–580 (1941).
[9] Schultz, J., and Bridges, C. B., *Am. Nat.*, **66**, 323–332 (1932).
[10] Houlahan, M. B., and Mitchell, H. K., PROC. NATL. ACAD. SCI., **33**, 223–229 (1947).
[11] Hannah, A. H., and Brown, S. W., Unpublished data.
[12] Knox, W. E., and Mehler, A. H., *J. Biol. Chem.*, **187**, 419–430 (1950).
[13] Mehler, A. H., and Knox, W. E., *Ibid.*, **187**, 431–438 (1950).
[14] Green, M. M., *Genetics*, **34**, 564–572 (1949).
[15] Dalgliesh, C. E., *Quart. Rev. Chem. Soc. (London)*, **5**, 227–244 (1951).
[16] Price, J. B., *U. Texas Publ.*, **4920**, 24–30 (1949).
[17] Sturtevant, A. H., and Novitski, E., *Genetics*, **26**, 517–541 (1941).
[18] Lewis, E. B., *Ibid.*, **30**, 137–166 (1945).
[19] Green, M. M., and Green, K. C., PROC. NATL. ACAD. SCI., **35**, 586–591 (1949).

THE PARACOMPACTNESS OF THE WEAK SIMPLICIAL COMPLEX

BY D. G. BOURGIN

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS

Communicated by W. H. Rodebush, January 25, 1952

The main purpose of this paper is to prove that a simplicial complex taken in the weak topology,[1] is a paracompact T_2 space.[2] Apart from this result, the method involves some new considerations which are expected to prove of wide application.

We assume throughout that $K = \bigcup K^n$, where K^n is the n dimensional skeleton. If U and V are collections of open sets we define the concept: V is a partial refinement of U, by requiring every element of V to be contained in some element of U. We write $St(s, U)$ for $Star\ (s, U)$. The symbol I or I_a (where a may be any script) *invariably* denotes the unit interval. We use both $\tilde{A}$ and $-A$ to denote complements so $B - A = B \cap \tilde{A}$. The topology giving the smaller open sets is the *finer* topology. In the weak (*or w*) topology for a simplicial complex K, open sets are those with relatively open intersections with each closed (Euclidean) simplex of K.

Denote the generic vertex of K by a and the totality by A. We use π in the dual sense of a finite subset of A constituting *the vertex set* of the open simplex $\sigma(\pi)$ of K, and for the *barycenter* of this simplex. The context will

make clear which sense is intended. We write $|\pi|$ for the number of points in π and in the sequel $|\pi_i| = i$. Let A_1 be the collection of vertices, π_i, of the barycentric derived complex K_1. The vertex set of a simplex of K is denoted by γ and the simplex by $E(\gamma)$. We write $E(\gamma) \epsilon \sigma(\pi)$ if $\pi' \epsilon \gamma \Longrightarrow \pi' \subset \pi$. The closed simplexes are $\overline{\sigma(\pi)}$ and $\overline{E(\gamma)}$. We order the simplexes $\{\sigma(\pi)\}$ by inclusion on π. Thus $\sigma(\pi) < \sigma(\pi') \Longleftrightarrow \pi \subset \pi'$. Similarly $E(\gamma) < E(\gamma') \Longleftrightarrow \gamma \subset \gamma'$.

1. The Barycentric Topology.—We shall consider K imbedded in the Cartesian product

$$\Pi = \Pi_A I_a. \tag{1}$$

This imbedding is a 1-1 correspondence between the points of K and those of a subset of Π. The points of Π are functions on A to I and are denoted by s or $s(a|A)$. The values of $\{s(a)\}$ for $s \epsilon K$ are the barycentric coordinates with respect to A. Hence if $s \epsilon \sigma(\pi)$ then $s(a|a \bar{\epsilon} \pi) = 0$ so $s(a|\pi)$ may be construed as barycentric coordinates with respect to π. Similarly we imbed K_1 in

$$_1\Pi = \Pi_{A_1} I_\pi, \tag{1.1}$$

with the generic points $t = t(\pi|A_1)$. Henceforth we understand by K and K_1 the corresponding complexes imbedded in Π and $_1\Pi$, respectively.

Define the 1-1 mapping i of K_1 onto K as follows: let t_0 be a point in the subcomplex $K_\pi \subset K_1$ consisting of the barycentric derived of $\sigma(\pi)$. Then t_0 is in a unique $E(\gamma)$ where $|\gamma| \leq |\pi|$. Let $i(t_0) = s_0$, where $s_0(a) = \sum_{a \epsilon \pi_j} t_0(\pi_j)/j$ with $t_0(\pi_l) = 0$ for $\pi_l \bar{\epsilon} \gamma$ and $s_0(a|a \bar{\epsilon} \pi) = 0$.

Topologize $_1\Pi$ by defining the neighborhood base $W(t_0) = \{\mathfrak{W}(t_0)\}$

$$\mathfrak{W}(t_0) = \Pi_{A_1} N(\pi, t_0), \tag{1.2}$$

where $N(\pi, t_0) = N(\pi, t_0(\pi))$ or merely $N(\pi)$, is an open interval in I_π. We topologize K by assigning the base, $\mathfrak{U}(s_0) = \{\mathfrak{U}(s_0)\}$,

$$\mathfrak{U}(s_0) = i(\mathfrak{W}(t_0) \cap K_1) \text{ or } i\mathfrak{W}(t_0), \tag{1.3}$$

understanding i is defined only on K_1. This topology is referred to as the barycentric topology.

Suppose $\sigma(\pi) \epsilon K$. Replace A_1 by π to get $_1\Pi_\pi$. λ_π is the inclusion map of $_1\Pi_\pi \rightarrow {}_1\Pi$ defined by putting in 0 values for the coordinates in $A_1 - \pi$. Then

$$i(i^{-1}(\sigma(\pi)) \cap \mathfrak{W}(t_0)) = i\lambda_\pi \Pi_\pi N(\pi', t_0).$$

This is denoted by $i_\pi \Pi_\pi N(\pi', t_0)$ and will be referred to as a $(\sigma(\pi))$ section neighborhood. If some point of $\mathfrak{W}(t_0) \cap K_1$ has a non-zero π' coordinate, we term π' an *essential* vertex and $N(\pi')$ an *essential* (coordinate) *neighborhood*.

Let $s_0 = i(t_0)$. Let $\sigma(s_0)$ and $E(t_0)$ be the unique simplexes in K and K_1 which contain s_0 and t_0. Let $\gamma(t_0)$ be the vertex set of $E(t_0)$.

$$Q = St(\sigma(s_0), K),$$
$$Q_1 = St(E(t_0), K_1). \tag{1.4}$$

If t_0 is not a vertex of K, let $\mathfrak{W}(t_0)$ be restricted by the requirement that for $\pi' \in \gamma(t_0)$, $N(\pi', t_0) \subset \{y \mid 0 < t_0(\pi') - \epsilon < y < t_0(\pi') + \epsilon < 1\}$. If t_0 is a vertex, say π^0, then require that $N(\pi^0) = \{y \mid 0 < a < y \leq 1\}$. For $\pi' \in \gamma(t_0)$ require $N(\pi') = \{y \mid 0 \leq y < \delta_{\pi'} < 1\}$. If $\mathfrak{W}(t_0)$ is restricted as described we refer to both $\mathfrak{W}(t_0)$ and $\mathfrak{U}(s_0) = i\mathfrak{W}(t_0)$ as *special* neighborhoods where t_0 may be replaced by any point $t \in \mathfrak{W}(t_0) \cap K_1$ and s_0 by $i(t)$. The utility of the special neighborhoods stems from their being entirely contained in Q or Q_1, respectively, and having all their non-vacuous sections connected. A *special section* neighborhood is a section of a special neighborhood.

2. *Equivalences.*—The following result, independently significant, is the basis for our demonstration of Theorem 3.

THEOREM 1. *The weak and the barycentric topologies are equivalent.*

Suppose G is open in the w topology. Let $s_0 \in G$. Order the *special neighborhoods* $\{\mathfrak{U}(s_0)\}$ *by* $\mathfrak{U}^1(s_0) < \mathfrak{U}^2(s_0)$ *if and only if* $\overline{\mathfrak{U}^1(s_0)} \cap \sigma(\pi) \subset G \cap \overline{\sigma(\pi)}$ *for* $\sigma(\pi) \in Q$ *implies* $N^1(\pi', t_0) = N^2(\pi', t_0)$ *for all subsets* π' *of* π. Accordingly $\mathfrak{U}^2(s_0) \cap \sigma(\pi) = \mathfrak{U}^1(s_0) \cap \sigma(\pi)$. If $\mathfrak{U}^1(s_0) < \mathfrak{U}^2(s_0)$ and $\mathfrak{U}^2(s_0) < \mathfrak{U}^1(s_0)$ we say $\mathfrak{U}^1(s_0)$ and $\mathfrak{U}^2(s_0)$ are equivalent. Let $\{\mathfrak{U}^\rho(s_0)\}$ be linearly ordered. Let $\sum = \{\sigma(\pi) \mid \sigma(\pi) \in Q, \mathfrak{U}^\rho(s_0) \cap \sigma(\pi) \subset G \cap \overline{\sigma(\pi)}\}$ for some ρ. Let Λ be the collection of all vertices in the barycentric subdivision of the simplexes of $\sum$. Let π' be such a vertex. Then $\sigma(\pi')$ is an element of $\sum$ and accordingly $N^\rho(\pi', t_0)$ is the same for all $\rho > \rho_0 = \rho_0(\pi')$ and may be designated by $N(\pi', t_0)$. Let $\mathfrak{W}(t_0) = \Pi_\Lambda N(\pi', t_0) \times \Pi_{\tilde{\Lambda}} N'(\pi', t_0)$ where $N'(\pi', t_0)$ is arbitrary subject to the restriction on special neighborhoods. Thus $\mathfrak{U}(s_0) = i\mathfrak{W}(t_0)$ is an upper bound for $\{U^\rho(s_0)\}$. Accordingly Zorn's lemma is applicable to assure the existence of a maximal special neighborhood, $\mathfrak{U}_m(s_0)$, according to our italicized ordering convention.

Suppose $\mathfrak{U}_m(s_0) \cap \overline{\sigma(\pi)}$ is *not* contained in $G \cap \overline{\sigma(\pi)}$ for say $\pi = \pi^*$, where $\sigma(\pi^*) \in Q$, and hence for all $\pi \supset \pi^*$. We may as well assume $\mathfrak{U}_m(s_0) \cap \sigma(\pi') \subset G \cap \overline{\sigma(\pi')}$, if π' satisfies $\sigma(s_0) < \sigma(\pi') < \sigma(\pi^*)$ (for otherwise we can replace π^* by one of its subsets and so attain the desired situation by a finite number of reductions). Let Γ indicate the collection of vertices in $K_1{}^*$. Since we are dealing with special neighborhoods, $N(\pi', t_0)$ may be replaced by 0 for $\pi' \in \Gamma$ without affecting $\mathfrak{U}_m(s_0) \cap \sigma(\pi^*)$. We assert we can redefine $N(\pi^*, t_0)$ as $N'(\pi^*, t_0) = \{x \mid 0 \leq x < \epsilon\}$ by taking ϵ small enough so that:

$$\overline{\sigma(\pi^*) \cap i(N'(\pi^*, t_0) \times \Pi_\Gamma^{\sim} 0_{\pi'} \times \Pi_{\Gamma - \pi^*} N(\pi', t_0))} \subset G^* \cap \overline{\sigma(\pi^*)}. \quad (1.5)$$

Indeed for $N(\pi^*, t_0)$ replaced by 0 we are reduced to the situation on the faces of $\sigma(\pi^*)$ which contain $\sigma(s_0)$ and by hypothesis the inclusion holds here. Since $\sigma(\pi^*)$ is a Euclidean simplex and $\bar{G} \cap \sigma(\pi^*)$ is closed, the distance from $i_\Gamma \Pi_\Gamma N(\pi', t_0) \cap \sigma(\pi^*)$ to $\bar{G} \cap \sigma(\pi^*)$ is positive. Hence an obvious compactness argument shows ϵ may be so chosen that (1.5) is satisfied. Evidently modifying $N(\pi^*)$ does not affect $\bigcup_m(s_0) \cap \sigma(\pi)$ for any π non $\supset \pi^*$. Accordingly the modified $\bigcup_m(s_0)$, denoted by $\bigcup_{m'}(s_0)$, satisfies $\bigcup_m(s_0) < \bigcup_{m'}(s_0)$ though $\bigcup_m(s_0)$ and $\bigcup_{m'}(s_0)$ are not equivalent. However, the existence of such a neighborhood $\bigcup_{m'}(s_0)$ is in contradiction with the maximality of $\bigcup_m(s_0)$. In short, $\overline{\bigcup_m(s_0) \cap \sigma(\pi)} \subset G \cap \sigma(\pi)$ for all $\sigma(\pi)$ in Q. Thus every weak neighborhood contains a barycentric neighborhood.

On the other hand, it is plain that every barycentric neighborhood is a weak neighborhood for the section neighborhoods are open in $\sigma(\pi)$.

3. Fineness of Topologies.—The natural topology[3] for K is given by the metric $d(s_0, s_1) = \left(\sum(s_0(a) - s_1(a))^2\right)^{1/2}$ where, evidently, only a finite number of terms in the sum are non-zero.

THEOREM 2. *The weak topology for K is finer than the natural topology.*

Evidently the weak topology is finer than that defined by the neighborhood basis with elements

$$\bigcup(s_0) = \Pi_A N(a, s_0) \cap K. \quad (3)$$

Let $G(s_0)$ be the sphere of radius δ about s_0 in the metric topology. Suppose $s_0 \in \sigma(\pi)$. Choose $N(a; s_0) = \left\{s(a) \middle| |s(a) - s_0(a)| < \delta^2/3|\pi|\right\}$ for $a \in \pi$. Any $\bigcup(s_0)$ so restricted lies inside $G(s_0)$. Indeed for $s \in U(s_0)$ one infers from $\sum_\pi s_0(a) = 1 = \sum s(a)$ and $\sum(s(a))^2 < \sum s(a)$ that

$$(d(s, s_0))^2 \leq 2\sum_\pi |s(a) - s_0(a)| < \delta^2.$$

Remark: The topology for K defined by (3) may be termed the "convex" topology. It is intermediate in fineness with the natural and the weak topologies, but is *not* equivalent to either in general.

4. The Main Result.—We present our key theorem.

THEOREM 3. *A simplicial complex with the weak topology is paracompact.*

It is necessary to show that every open cover U of the simplicial complex K has an open Δ refinement.[4] In view of Theorem 1 we can understand $\bigcup$ to be given by (1.3) and indeed to be a special neighborhood of some point s. Throughout the proof, k will invariably be an ordering script taking on the values 0, 1, 2 and 3.

As a first step in an induction argument choose special neighborhoods $\left\{{}^k V(a)\right\}$ for each vertex $a = i\pi_1$ to satisfy

$$a \subset {}^3V_1(\pi_1) \subset {}^3\bar{V}_1(\pi_1) \subset \ldots \subset {}^0V_1(\pi_1) \subset \bigcup \quad (4)$$

for some $U \in U$. Since these are special neighborhoods it is plain that ${}^0V_1(\pi_1) \cap {}^0V_1(\pi_1') = \phi$ for $\pi_1 \neq \pi'_1$. The existence of $\{{}^kV_1(\pi_1)\}$ is easily established. Fix a vertex, $a = i\pi_1$, corresponding to $s_0 = i(t_0)$ with $U(s_0)$ a special neighborhood given by (1.3). Thus for $\pi \neq \pi_1$, $N(\pi, t_0) = \{t(\pi) | 0 \leq t(\pi) < 1 - \delta_\pi < 1\}$ while, for $\pi = \pi_1$, $N(\pi_1, t) = \{t(\pi_1) | 0 < \delta_{\pi_1} < t(\pi_1) \leq 1\}$. Define ${}^kN(\pi, t_0)$ by replacing δ_π by $\delta_\pi(1 + {}^k\rho_\pi)$ subject to $(1 - \delta_\pi)/\delta_\pi > {}^3\rho(\pi) > \ldots > {}^0\rho(\pi) > 0$. The neighborhoods ${}^kV_1(\pi_1) = i\Pi^kN(\pi, t_0)$ satisfy (4). We write

$$ {}^kV(\pi_1) = \{{}^kV_1(\pi_1)\}, \quad V(\pi_1) = \{{}^2V_1(\pi_1), {}^1V_1(\pi_1) - {}^3\bar{V}_1(\pi_1)\}, $$
$$ \bar{V}(\pi_1) = \{{}^2\bar{V}_1(\pi_1), {}^1\bar{V}_1(\pi_1) - {}^3V_1(\pi_1)\} \quad (4.1) $$

It is trivial that ${}^kV(\pi_1)$, $V(\pi_1)$, $\bar{V}(\pi_1)$ constitute partial Δ refinements of U.

We develop the induction argument. Let $\pi = \pi_n = \{a_i | i = 1, \ldots n\}$. Let π_r, π_r' be subsets containing r elements of π, $1 \leq r \leq n - 1$. We shall often write ${}^kV_j(\pi_r)$ for ${}^kV_{j(\pi_r)}(\pi_r)$ and ${}^{13}V_j(\pi_r)$ for ${}^1V_j(\pi_r) - {}^3\bar{V}_j(\pi_r)$. Thus ${}^{13}\bar{V}_j(\pi_r) = {}^1\bar{V}_j(\pi_r) - {}^3V_j(\pi_r)$.

Assume there are covers $V(\pi_{n-1})$ and $\bar{V}(\pi_{n-1})$ of $\sigma(\pi_{n-1})$ for each $\pi_{n-1} \subset \pi$ which are partial Δ refinements of U with the members of $\bar{V}(\pi_{n-1})$ the closures of those of the open cover $V(\pi_{n-1})$, where for $r = 1, \ldots, n - 1$,

$$ V(\pi_r) = \mathbf{U}_{\pi_{r-1} \subset \pi_r} V(\pi_{r-1}) \cup \mathbf{U}^2V_j(\pi_r) \cup {}^{13}V_j(\pi_r), $$

where $j = j(\pi_r)$ takes on the values 1 through $M(\pi_r)$ comprising the set $M(\pi_r)$ and $M(\pi_1) = 1$, and ${}^kV_j(\pi_r)$ is a special neighborhood. We assume also

$$ {}^3V_j(\pi_r) \subset {}^3\bar{V}_j(\pi_r) \subset \ldots {}^0V_j(\pi_r) \quad (4.3) $$

and

$$ {}^kV(\pi_{r+1}) = \mathbf{U}_{\pi_r \subset \pi_{r+1}} V(\pi_r) \cup \mathbf{U}_{l \in M(\pi_{r+1})} {}^kV_l(\pi_{r+1}) \quad (4.4) $$

is a partial Δ refinement of U which covers $\sigma(\pi_{r+1})$ for each $r = 1, \ldots n - 2$. Observe that if $\pi_r \neq \pi_r'$ then *the sets* $\{{}^0\bar{V}_l(\pi_r) | l \in M(\pi_r)\}$ *are disjunct from the sets* $\{{}^0V_j(\pi_r') | j \in M(\pi_r')\}$ *since these are special neighborhoods.* This property is designated by S. The assumptions in this paragraph constitute the induction assumptions. The guiding idea in the ensuing construction of $\bar{V}(\pi)$ is to ensure that *for every* s *there is an* $s_\pi \in \sigma(\pi)$ *for which* $St(s, \bar{V}(\pi)) \subset St(s_\pi, \bar{V}(\pi)) \subset U$.

Let $T(\pi)$ be the set $\{l(\pi_m) | l(\pi_m) \in M(\pi_m), \pi_m \subset \pi, m = 1, \ldots n - 1\}$ and let $L(\pi) = \sigma(\pi) - \mathbf{U}_{T(\pi)} {}^1V_l(\pi_m)$. Consider the cover of $L(\pi)$ by $U' = U\Lambda \tilde{\mathbf{U}}_{\tau(\pi)} {}^2V_l(\pi_m)$. Replace U' by U'' a refinement whose elements are special neighborhoods. The compactness of $L(\pi)$ implies there is a finite subcover by $\{U''_j(\pi)\}$. Write $U''_j(\pi) \cap \sigma(\pi) = P^j(\pi)$. Let $\mu = \{u(\pi_i) | i = 1, \ldots n - 1\}$, $\rho = \{\rho(\pi_i) | i = 1, \ldots n - 1\}$ where $\mu(\pi_i) \cup \rho(\pi_i) = M(\pi_i)$. Define

$$Z_\rho^\mu = \bigcap_{j \in \rho(\pi_r),\ r=1,\ \ldots\ n-1} {}^0V_j(\pi_r) \ \cap\ \tilde{\bigcup}_{l \in \mu(\pi_m),\ m=1,\ \ldots\ n-1} {}^1\bar{V}_l(\pi_m) \quad (4.5)$$

where $\rho = 0$ indicates the null set. Some sets, Z_ρ^μ, may be empty. Let $x = (j, \rho, \mu) = (j(x), \rho(x), \mu(x))$. Define

$$P_x(\pi) = P^j(\pi) \ \cap\ Z_\rho^\mu \quad (4.6)$$

Designate $\{P_x(\pi)\}$ by $P(\pi)$ a cover of $L(\pi)$. The union of $P_x(\pi)$ sets for fixed j is $P^j(\pi)$. Since $\bigcup P_x$ is a subset of a Euclidean $n-1$ dimensional space, there is a * (star) refinement,[3] $Q(\pi)$, of $P(\pi)$ with elements special section neighborhoods. Furthermore we may require that the closure of the elements in $Q(\pi)$ is still a * refinement of $P(\pi)$. We need only a finite number of sets in $Q(\pi)$ to cover $L(\pi)$. Denote them by $H(\pi) = \{H_l(\pi)\}$. Thus $H_l(\pi)$ is the section neighborhood $H_l(\pi) = i_\pi \Pi_\pi N(\pi')$. *We assume furthermore that the essential neighborhoods defining $H_l(\pi)$ are minimal* in the sense that if $H_l(\pi)$ is also $i_\pi \Pi_\pi N'(\pi')$ then $N(\pi') \subset N'(\pi')$.

We assert we can replace $H(\pi)$ by minimal special section neighborhoods, ${}^kH(\pi) = \{{}^kH_l(\pi)\}$, satisfying the condition that ${}^3H_l(\pi)$ cover $L(\pi)$ and

$$\left.\begin{array}{lll} {}^3H_l(\pi) \subset {}^3\bar{H}_l(\pi) \subset & & \subset {}^0H_l(\pi) \subset H_l(\pi); \quad \text{(a)} \\ {}^kH_l(\pi) \cap {}^1V_m(\pi_r) = \phi & \Longrightarrow & {}^kV_l(\pi) \cap {}^1V_m(\pi_r) = \phi; \quad \text{(b)} \\ \bar{H}_l(\pi) \cap {}^1V_m(\pi_r) \neq \phi & \Longrightarrow & {}^3\bar{V}_l(\pi) \cap {}^1V_m(\pi_r) \neq \phi; \quad \text{(c)} \\ H_l(\pi) \cap {}^{13}\bar{V}_m(\pi_r) \neq \phi & \Longrightarrow & {}^3V_l(\pi) \cap {}^{13}\bar{V}_m(\pi_r) \neq \phi. \quad \text{(d)} \end{array}\right\} \quad (4.7)$$

we write (4.7e) and (4.7f) when the bars are dropped in (4.7c) and (4.7d).

Thus we obtain ${}^kH_l(\pi)$ by replacing $N(\pi')$ in the definition of $H_l(\pi)$ by ${}^k\rho(\pi')N(\pi')$ where the symbolism denotes a shrinkage of $N(\pi')$, defined by replacing δ_i by $\delta_i\ (1 + \rho)$, $i = 1, 2$, for the lower, δ_1, and upper, $1 - \delta_2$, end points of the neighborhood $N(\pi')$. (Observe $\delta_i = 0$ is preserved.) Here

$$\frac{1 - \delta_1 - \delta_2}{2} > {}^3\rho(\pi') > \ \ldots\ > {}^0\rho(\pi') > 0. \quad (4.8)$$

The demonstration that ${}^k\rho(\pi')$ exists to validate (4.7b) is simple but typical for later situations and is therefore labeled *Argument A*. Thus the negation of (4.7b) implies there is a boundary point $\tau_0 = \{\tau_0(\pi_i) | \pi_i \subset \pi\}$ of $\Pi_\pi N(\pi_i) \times N'(\pi_i) \cap \bar{K}_\pi$, where $N(\pi_i)$ and $N'(\pi_i)$ indicate the neighborhoods in ${}^kH_l(\pi)$ and ${}^1V_m(\pi_r)$, respectively, for fixed l, k and m. Hence $\tau_0(\pi_i)$ is a boundary point of $N(\pi_i) \times N'(\pi_i)$. For $\pi_i \subset \pi$ let $N(\pi_i) = \{\tau(\pi_i) | a_i < \tau(\pi_i) < b_i\}$ and $N'(\pi_i) = \{\tau(\pi_i) | A_i < \tau(\pi_i) < B_i\}$, where $<$ can be replaced by $\leq$ if the lower bound is 0 or the upper bound is 1. Accordingly $\tau_0(\pi_i)$ is one of the values a_i, b_i, A_i, B_i. Suppose for all i, $\tau_0(\pi_i)$ is neither a_i nor b_i. Then $i_\pi \tau_0$ is a boundary point of ${}^1V_m(\pi_r) \cap \sigma(\pi)$. Moreover $i_\pi \tau_0 \in H_l(\pi)$. Hence every neighborhood of $i_\pi \tau_0$ contains points of ${}^1V_m(\pi_r)$ and so $H_l(\pi) \cap {}^1V_m(\pi_r) \neq \phi$ contrary to our assumption. Accordingly the set,

α, of i's for which $\tau_0(\pi_i) = a_i$ or b_i is not empty. For $i \in \alpha$ the minimality hypothesis guarantees that *any* change in ${}^k\rho^l(\pi_i)$ effects changes in ${}^kH_l(\pi)$ that eliminate τ_0 as a common boundary point for ${}^kH_l(\pi)$ and ${}^1V(\pi_r)$ and changes in at most $|\pi|\ {}^k\rho^l(\pi_i)$'s eliminate all common boundary points. Since there are only a finite number of ${}^1V_m(\pi_r)$ sets, ${}^k\rho^l(\pi_i)$ need merely stay away from a *finite* number of values.

In view of (4.7b) the closures may be dropped in (4.7c) and (4.7d) and (4.7e) ... (4.7f) follow easily now for ${}^3\rho^l(\pi)$ sufficiently small.

Define the collections

$$\begin{aligned} {}^kC(\pi) &= \{ V(\pi_r), {}^kH_l(\pi) \}, \\ D(\pi) &= \{ {}^{13}\bar{V}_l(\pi_r) \cap \sigma(\pi), \quad {}^0H_l(\pi) \}, \end{aligned} \tag{4.9}$$

where $r = 1, \ldots n-1$ and l as usual stands for $l(\pi_r)$. Refer to $St(s, D(\pi)) = \{ W_b, {}^0H_l | b \in \beta, l \in \lambda \}$ as the *intersection set* $[\beta, \lambda]$ where W_b is one of the elements ${}^{13}\bar{V}_l(\pi_r) \cap \sigma(\pi)$. Here *either β or λ may be the empty set.* Write $W_b = \Pi_\pi \bar{N}^b(\pi')$, $\bar{N}^\beta(\pi') = \bigcap_\beta \bar{N}^b(\pi')$. Also ${}^0H_l(\pi) = i_\pi \Pi_\pi N_l(\pi')$ and $N^\lambda(\pi') = \bigcap_\lambda N^l(\pi')$. Write $N^{\beta,\lambda}(\pi') = \bar{N}^\beta(\pi') \cap N^\lambda(\pi')$. We say $[\beta, \lambda]$ *admits* ${}^0H_m(\pi)$ if for all $\pi' \subset \pi$, $N^m(\pi') \cap N^{\beta,\lambda}(\pi') \neq \phi$ yet ${}^0H_m(\pi) \cap [\beta, \lambda] = \phi$.

Let $M(\beta, \lambda, m)$ be the supremum of $\sum_\pi (n_i | n_i \in N^m(\pi') \cap N^{\beta,\lambda}(\pi'))$. If $M(\beta, \lambda, m) = 1$ we can replace ${}^0H_m(\pi)$ by a slightly smaller minimal neighborhood ${H_m}'(\pi)$, so that $M(\beta, \lambda, m)$ is now inferior to 1, such that all intersection relations are unaffected by replacing $H_m(\pi)$ by ${H_m}'(\pi)$. In particular the intersection sets remain intersection sets with this replacement. We can require moreover that ${}^0C(\pi)$ remain a cover and (4.7) be maintained. The demonstration pivots on Argument A.

We write ${}^0J_l(\pi)$ for ${}^0H_l(\pi)$ or for ${}^0{H_l}'(\pi)$ if this last set has been introduced and have therefore ${}^0J(\pi)$. Evidently replacing ${}^0H_l(\pi)$ by ${}^0J_l(\pi)$ in an intersection set cannot increase $M(\beta, \lambda, m)$ so use of ${}^0J(\pi)$ in place of ${}^0H(\pi)$ guarantees the constants $M(\beta, \lambda, m)$ are all less than 1 and $[\beta, \lambda]$, *except for possible* ${}^2V_l(\pi_r)$ *sets,* is $St(s, {}^0\bar{C}(\pi))$ *for some* $s \in \bar{\sigma}(\pi)$.

Consider now the collection $\{ W_g | \Gamma \}$ satisfying (*a*) $\bigcap_\Gamma W_g = \phi$, (*b*) $\bigcap_\Gamma \bar{N}^g(\pi' | \pi) \neq \phi$ and for some m, ${}^0J_m(\pi) \cap W_g \neq \phi$, for all $g \in \Gamma$. We say ${}^0J_m(\pi)$ is admitted by the *quasi intersection set* $[\Gamma]$. Let $M(m, \Gamma) =$ max. $\sum_\pi n(\pi')$, for all $n(\pi' | \pi) \in \bar{N}^m(\pi') \cap \bar{N}^g(\pi')$. Then $M(m, \Gamma)$ is inferior to 1. Let $M(m)$ be max. $(M(\beta, \lambda, m), M(m, \Gamma))$ where comparison is made for all intersection and quasi-intersection sets which admit ${}^0J_m(\pi)$.

We wish now to replace ${}^kJ_l(\pi)$ by ${}^kV_l(\pi)$ to get the collection ${}^kV(\pi)$. We write ${}^0N^m(\pi' | \pi)$ for the essential neighborhoods making up ${}^0J_m(\pi)$. We require ${}^0N^m(\pi')$, for π' non $\subset \pi$, to satisfy a *first condition*

$${}^0N^m(\pi') \subset \{ t(\pi') | 0 \leq t(\pi') < (1 - M(m)) 3^{-(|\pi|+1)} \}. \tag{4.10}$$

Suppose ${}^0J_m(\pi)$ is contained in $P_x(\pi)$ for all x in the set $X(m)$ and for no x

outside the set. Then, in view of (4.6),

$$i_\pi \Pi_\pi N^m(\pi') \subset \bigcap_{X(m)} P_x(\pi),$$

$$P_x(\pi) \subset i(\bigcap_{r=1}^{r=n-1} \bigcap_{l(\pi_r) \in \rho(x)} {}^0V_l(\pi_r) \cap \bigcup^{''j(x)} \cap \overline{\sigma(\pi)}. \quad (4.11)$$

Suppose the constituent neighborhoods of ${}^0V_l(\pi_r)$ and $\bigcup^{'' j(x)}$ are denoted by $N^l(\pi')$ and $N^{j(x)}(\pi')$, respectively. Write $\rho(x) = \{\rho(\pi_r, x)\}$,

$$N(\pi', \rho(x)) = \bigcap_{l \in \rho(x)} N^l(\pi'),$$

then

$$P_x(\pi) \subset i(\Pi_{A_l} N(\pi', \rho(x)) \cap N^{''j(x)}(\pi') \cap \bar{K}_\pi).$$

If $\rho(\pi_r, x) = 0$ for all π_r we drop the $N(\pi', \rho(x))$ term. We state a *second condition*,

$${}^0N^m(\pi') \subset \bigcap_{x \in X(m)} N(\pi', \rho(x)) \cap N^{''j(x)}(\pi'), \pi' \text{ non } \subset \pi \quad (4.12)$$

subject to $N^m(\pi')$ is a neighborhood of 0 in $I_{\pi'}$. For $\pi' \subset \pi$ a similar inclusion relation holds automatically as a consequence of (4.11). Hence we define ${}^0V_m(\pi)$ with ${}^0V_m(\pi) \cap \bar{\sigma}(\pi) = {}^0J_m(\pi)$ by

$${}^0V_m(\pi) = i(({\lambda_\pi}^{-1}\, {}^0J_m(\pi)) \times \Pi_{\pi' \text{ non } \subset \pi} N^m(\pi')),$$

subject to conditions (4.10) and (4.12). Next choose the positive constants ${}^k\rho^m(\pi)$, $k \neq 0$, to satisfy the conditions (4.8) and so that (4.7) is satisfied and we have

$${}^kV_m(\pi) = i(({\lambda_\pi}^{-1}\, {}^kJ_m(\pi)) \times \Pi_{\pi' \text{ non } \subset \pi} {}^k\rho^m(\pi') N^m(\pi')). \quad (4.13)$$

we remark

$${}^0J_l(\pi) \cap {}^{13}\Gamma_m(\pi_r) = \phi \Longrightarrow {}^0\bar{V}_l(\pi) \cap {}^{13}\Gamma_m(\pi_r) = \phi. \quad (4.14)$$

We state as our Minor Thesis that *if* $s_0 \in \bar{\sigma}(\pi)$ *then* $St(s_0, {}^0\bar{V}(\pi)) \subset \bigcup$. Suppose then $St(s_0, {}^0\bar{V}(\pi)) = G_1 \cup G_2 \cup G_3$ where $G_1 = \{{}^0\bar{V}_l(\pi) | \Lambda\}$, $G_2 = \{{}^{13}\bar{V}_b(\pi_r) | b \in \beta\}$ and $G_3 = \{{}^2\bar{V}_g(\pi_r) | g \in \Gamma\}$. Suppose Λ is empty but either or both of β and Γ are non-empty. Let $R \leq n - 1$ be the maximum value of r for any neighborhood in $G_2 \cup G_3$. Then $s_0 \in {}^0\bar{V}(\pi_R)$ where π_R is unique by S. This case is settled then by appeal to our induction assumptions. Suppose Λ and β are non-empty. Then Γ must be empty by (4.6). Let $R \leq n - 1$ be the maximum value of r for any set in G_2. Recall all members of $G_1 \cap \overline{\sigma(\pi)}$ must be in a common $P_x(\pi)$. Since P_x intersects ${}^1V_m(\pi_R)$ it must be in ${}^0V_m(\pi_R)$ according to (4.5) and (4.6). Then (4.11), (4.12) and (4.13) guarantee $\{{}^0\bar{V}_l(\pi) | \Lambda\} \subset {}^0V_m(\pi_R)$. Hence s_0 is covered by ${}^0\bar{V}(\pi_R)$ Hence by the induction assumptions and S we have $\bigcup \supset St(s_0, {}^0\bar{V}(\pi_R)) = St(s_0, {}^0\bar{V}(\pi))$. Suppose now Λ alone is non-empty. Then $P_x(\pi) \supset G_1 \cap \overline{\sigma(\pi)}$. Hence by (4.12) and (4.13), $\bigcup^{''j(x)} \supset St(s_0, {}^0\bar{V}(\pi))$.

We now establish ${}^0\bar{\Gamma}(\pi)$ is a cover of $L(\pi)$ *partially* Δ *refining* U. If $St(s_0, {}^0\bar{\Gamma}(\pi))$ does not contain any sets $\{{}^0\bar{\Gamma}_l(\pi)\}$ then let $R \leq n-1$ be the maximum value of r for which s_0 is in a set ${}^0\bar{\Gamma}_m(\pi_r)$. Then π_R is unique, since these are special neighborhoods and $St(s_0, {}^0\bar{\Gamma}(\pi)) = St(s_0, {}^0\bar{\Gamma}(\pi_R))$ which is contained in some $\cup$ by the induction assumption. In the sequel denote the sets ${}^{13}V_l(\pi_r)$ by ${}^{13}V_b$ and suppose, next, $St(s_0, {}^0\bar{\Gamma}(\pi)) = F_1 \cup F_2$ where $F_1 = \{{}^0\bar{\Gamma}_l(\pi) | \Lambda\}$, and $F_2 = \{{}^{13}\bar{\Gamma}_b | \beta\}$ with *neither* β *nor* Λ *a null set.* Let G_1 and G_2 be the collections obtained from F_1 and F_2, respectively, by intersecting their members with $\bar{\sigma}(\pi)$. It is impossible that no member of ${}^0\bar{J}(\pi)$ intersects every element of G_2 for by (4.14) this would imply F_1 is vacuous. Moreover G_2 cannot be a quasi-intersection set admitting any ${}^0J_l(\pi)$. Indeed if the admissible ${}^0J_l(\pi)$ is in G_1 then by (4.10), ${}^0\bar{\Gamma}_l(\pi)$ cannot be in F_1. If ${}^0J_l(\pi)$ is not in G_1 then for some b, ${}^0J_l(\pi) \cap {}^{13}\bar{\Gamma}_b = \phi$ and (4.14) again disposes of this possibility. We have just showed the sets of G_2 have a common point. If ${}^0J_m(\pi)$ is admitted by an intersection set including G_2 then ${}^0\bar{\Gamma}_m(\pi)$ cannot be in F_1 in view of (4.10). Hence we have shown $G_1 \cup G_2$ is part of an intersection set and so there is a point s_π in $\bar{\sigma}(\pi)$ for which $St(s_\pi, {}^0\bar{C}(\pi)) \supset G_1 \cup G_2$. Hence, $St(s_\pi, {}^0\bar{\Gamma}(\pi)) \supset F_1 \cup F_2 = St(s_0, {}^0\bar{\Gamma}(\pi))$ and we have only to refer to our Minor Thesis: $St(s_\pi, {}^0\bar{\Gamma}(\pi)) \subset \cup$.

If β is the null set, F_2 and G_2 are empty. Then G_1 is part of an intersection set, in view of (4.10). Then for some s_π we have again $St(s_\pi, {}^0\bar{C}(\pi) \supset G_1$ and by reason of our Minor Thesis $\cup \supset St(s_\pi, {}^0\bar{\Gamma}(\pi)) \subset F_1 = St(s_0, {}^0\bar{\Gamma}(\pi))$. For instance if s_π is not in any ${}^{13}V_m(\pi_r)$, $G_1 \subset P_x$ and $\cup^{\prime\prime j(x)}$ serves as our $\cup$.

We have shown then that ${}^0\bar{\Gamma}(\pi)$ is a cover of $\bar{\sigma}(\pi)$ partially Δ refining U. Furthermore since $\{{}^3J_l(\pi)\}$ is a cover of $L(\pi)$ it follows that both $V(\pi)$ and $\bar{\Gamma}(\pi)$ are covers of $\bar{\sigma}(\pi)$ partially Δ refining U.

The open cover of K^0 partially Δ refining U is $\{{}^0V(a) | A\}$. The choice of ${}^kV_l(\pi)$ depends only on the neighborhoods $\{{}^kV_m(\pi_r) | \pi_r \subset \pi\}$ and is independent of the sets $\{{}^kV_m(\pi_q)\}$ for π_q non $\subset \pi$. Hence our induction has actually shown that the existence of an open cover of K^{n-1} whose closure partially Δ refines U implies the existence of an open cover of K^n whose closure partially Δ refines U where the cover for K^n is obtained from that for K^{n-1} by *adding some new open sets.* Hence the union of all these covers for all n is an open cover of K which Δ refines U. The theorem is therefore established.

[1] Whitehead, J. H. C., "Combinatorial Homotopy I," *Bull. Am. Math. Soc.*, **55**, 213–245 (1949).

[2] Dieudonne, J., "Une generalisation des espaces compact," *J. Math. Pure Appl.*, **23**, 65–76 (1944).

[3] Lefschetz, S., *Topics in Topology*, Princeton University Press, 1942, pp. 1–137.

[4] Stone, A. H., "Paracompactness and Product Spaces," *Bull. Am. Math. Soc.*, **54**, 977–982 (1948).

SOME PROBLEMS INVOLVING PRIMITIVE ROOTS IN A FINITE FIELD

By L. Carlitz

Department of Mathematics, Duke University

Communicated by H. S. Vandiver, February 18, 1952

Introduction.—In this note we state a number of results concerning primitive roots in a finite field. Detailed proofs will appear elsewhere. It is convenient to consider the results under three heads.

For §1 we require certain notions introduced by Ore[5]. For $\gamma \in GF(p^{nm})$, $\gamma \neq 0$, let

$$a(x) = \sum_{r=0}^{k} a_r x^{p^{nr}} \qquad (a_r \in GF(p^n),\ a_k \neq 0),$$

be a "linear" polynomial of minimal order k such that $a(\gamma) = 0$; then γ is said to *belong* to $a(x)$. If $A = A(x) = \sum_{0}^{k} a_r x^r$, then $a(x)$ *corresponds* to $A(x)$. In particular if γ belongs to $x^{p^{nm}} - x$, it is a primitive root (of the second kind); an ordinary primitive root may be called one of the first kind. It is natural to ask whether there exist numbers that are simultaneously primitive roots of the first and second kind, more generally whether there exist numbers belonging simultaneously to preassigned e and $a(x)$, where $e \mid p^{nm} - 1$ and $A(x) \mid x^m - 1$. We show that the answer is affirmative provided e deg $a(x)$ and p^n are sufficiently large.

In §2 we extend some theorems of Davenport[1] on the distribution of primitive roots.

In §3 we consider first the number of solutions $\beta_1, \ldots, \beta_r$ of

$$\alpha = \alpha_1\beta_1 + \ldots + \alpha_r\beta_r, \tag{1}$$

where $\alpha_i \neq 0$ and β_i belongs to the exponent e_i. Next we consider the number of solutions of

$$\alpha = \alpha_1\beta_1 + \ldots + \alpha_r\beta_r + \delta_1\xi_1^{k_1} + \ldots + \delta_s\xi_s^{k_s}, \tag{2}$$

where $\alpha_i\delta_i \neq 0$, ξ_i arbitrary, $k_i \mid p^n - 1$. Asymptotic results are obtained for both problems. However, for the number of solutions in the special case

$$\alpha = \gamma\beta + \delta_1\xi_1^2 + \ldots + \delta_{2s}\xi_{2s}^2 \qquad (p > 2), \tag{3}$$

a simple exact formula is obtained.

The starting point for most of the theorems of this paper is contained in Lemmas 1 and 2 below; these lemmas are suggested by the well-known proof of Vinogradoff's theorem on the smallest primitive root of a prime

(see ref. 4, p. 178). The method of proof for the most part involves manipulation of character sums; compare in particular Davenport and Hasse[2] and Hua and Vandiver.[3]

1. Primitive roots of the first and second kind. Let $\theta, \theta^{p^n}, \ldots, \theta^{p^{n(m-1)}}$ be a normal basis of $GF(p^{nm})$ relative to $GF(p^n)$. For $\alpha \in GF(p^{nm})$ put

$$\alpha = a_0\theta + a_1\theta^{p^n} + \ldots + a_{m-1}\theta^{p^{n(m-1)}} \qquad (a_r \in GF(p^n)).$$

Then we define

$$\lambda(\alpha) = \lambda_c(\alpha) = E(a_0c_0 + \ldots + a_{m-1}c_{m-1}),$$

where $E(a) = e^{2\pi i t(a)/p}$, $t(a) = a + a^p + \ldots + a^{p^{n-1}}$, and $c_0, \ldots, c_{m-1}$ are arbitrary numbers of $GF(p^n)$; in particular the function $\lambda_0(\alpha) \equiv 1$ corresponds to $c_0 = \ldots = c_{m-1} = 0$.

We next define

$$\Omega(\alpha) = \frac{1}{|B|} \sum_{D|A} \frac{\mu(D)}{|D|} \sum \lambda(DB)\,\lambda(\alpha) \qquad (|B| = p^{n \deg B}),$$

where $AB = x^m - 1$ and the inner sum is restricted to the λ's such that $\lambda(h(\gamma)) = \lambda_0$, where $h(x)$ corresponds to $H(x) = D(x)B(x)$.

LEMMA 1. *If α belongs to $a(x)$, then $\Omega(\alpha) = 1$; for all other α, $\Omega(\alpha) = 0$.*

We shall also require the function

$$\omega(\alpha) = \frac{1}{f} \sum_{d|e} \frac{\mu(d)}{d} \sum_{\chi^{df} = \chi_0} \chi(\alpha),$$

where $ef = p^n - 1$ and $\chi(\alpha)$ denotes a multiplicative character of $GF(p^{nm})$.

LEMMA 2. *If α belongs to the exponent e, then $\omega(\alpha) = 1$; for all other α, $\omega(\alpha) = 0$.*

Making use of Lemmas 1 and 2 we prove

THEOREM 1. *Let $N(e, a(x))$ denote the number of numbers of $GF(p^{nm})$ which simultaneously belong to e and $a(x)$. Then*

$$N(e, a(x)) = \frac{\phi(e)\Phi(A)}{p^{nm}} + O(p^{nm(1/2+\epsilon)}) \qquad (p^{nm} \to \infty).$$

Here $\phi(e)$ is the ordinary ϕ-function, while $\Phi(A)$ is the corresponding function for $GF[p^n, x]$.

As a particular case of Theorem 1 we have

THEOREM 2. *Let N' denote the number of numbers of $GF(p^{nm})$ that are simultaneously primitive roots of the first and second kind. Then*

$$N' = \frac{\phi(p^{nm}-1)\Phi(x^m-1)}{p^{nm}} + O(p^{nm(1/2+\epsilon)}).$$

It will now be convenient to use the concrete representation of $GF[p^n, x]/$

$P(x)$ for the $GF(p^{nm})$, where $P(x)$ is an irreducible polynomial $\epsilon GF[p^n, x]$ of degree m.

THEOREM 3. *Put* $P(x) = x^m + c_1x^{m-1} + \ldots + c_m$, $c_j \epsilon GF(p^n)$. *Then a necessary and sufficient condition that*

$$a(R) = R + R^{p^n} + \ldots + R^{p^{n(m-1)}} \equiv 0 \pmod{P}$$

for all $R \epsilon GF[p^n, x]$ *of degree* $\leq r$ *is furnished by*

$$p \mid m; \quad c_s = 0 \text{ for } 1 \leq s \leq r,\ p + s.$$

Using Theorem 3, we next prove

THEOREM 4. *For given* p^n, $r \geq 0$, *there exist infinitely many irreducibles* $P \epsilon GF[p^n, x]$ *such that* $a(R) = R + R^{p^n} + \ldots + R^{p^{n(m-1)}} \equiv 0 \pmod{P}$ *for all* R *of degree* $\leq r$, *where* $m = \text{deg. } P$.

In particular it follows that no polynomial R of degree $\leq R$ can be a primitive root of the second kind (mod P).

THEOREM 5. *Let* $N_r(e, a(x))$ *denote the number of polynomials of degree* $< r$ *which simultaneously belong to* e *and* $a(x) \pmod{P}$; *let* $N_r(a(x))$ *denote the number that belong to* $a(x)$. *Then*

$$N_r(a, e(x)) = \frac{\phi(e)}{p^{nm} - 1} N_r(a(x)) + O(p^{nm(1/2 + \epsilon)}).$$

Note that we are unable to find an estimate for $N_r(a(x))$.

2. Davenport[1] proved that if the prime p is sufficiently large, and θ is a generating element of $GF(p^n)$ then there exists an integer a such that $\theta - a$ is a primitive root of the field. On the other hand, he proved that for given p there exist fields $GF(p^n)$ and generating elements θ such that no number of the form $a\theta + b$ is a primitive root. We shall now extend these results.

THEOREM 6. *Let* $N(e)$ *denote the number of numbers* $\theta + a$ *that belong to the exponent* e. *Then*

$$N(e) = \frac{\phi(e)}{p^n - 1} p + O\left(p^{1 - \frac{1}{2(n+1)} + \epsilon}\right) \qquad (p \to \infty).$$

Using a theorem of A. Weil[6] we prove the more precise

THEOREM 7. *For* $r \geq 1$, *let* $N_r(e)$ *denote the number of numbers* $\alpha = \theta^r + a_1\theta^{r-1} + \ldots + a_r$, $a_i \epsilon GF(p)$, *that belong to the exponent* e. *Then*

$$N_r(e) = \frac{\phi(e)}{p^n - 1} p^r + O(p^{1/2r + \epsilon}) \qquad (p \to \infty).$$

We have assumed n, r fixed while $p \to \infty$. The next theorem is valid for all p.

THEOREM 8. *For all* p *and* $1 \leq r \leq n$, *let* $N_r^*(e)$ *denote the number of numbers*

$$\alpha = \alpha_0 + c_1\alpha_1 + \ldots + c_r\alpha_r \qquad (c_i \in GF(p))$$

that belong to the exponent e, where $\alpha_1, \ldots, \alpha_r$ are linearly independent relative to $GF(p)$. Then

$$N_r^*(e) = \frac{\phi(e)}{p^n - 1}\, p^r + O(p^{n(1/2+\epsilon)}).$$

In particular if $r > {}^1/_2 n$ and $e \geq p^c$, $c > {}^3/_2 n - r$, then

$$N_r^*(e) \sim \frac{\phi(e)}{p^n - 1}\, p^r \qquad (p^n \to \infty).$$

The next theorem is an extension of (ref. 1, Th. 2).

THEOREM 9. *For fixed p^n, $r \geq 1$, there exist infinitely many irreducible polynomials $P(x) \in GF[p^n, x]$ such that*

$$A(p^{nm} - 1)/(p^n - 1) \equiv 1 \pmod{P} \qquad (m = \deg P).$$

In particular no such A is a primitive root (mod P). This theorem may be compared with theorem 4 above.

The final theorem of this section is concerned with integers (mod p).

THEOREM 10. *Let $e \mid p - 1$, $r \geq 1$, $s \geq 0$. Let $N(e, r) = N(e, r, s)$ denote the number of integers t, $s < t \leq s + r$, which belong to the exponent e (mod p). Then*

$$N(e, r) = \frac{\phi(e)}{p - 1}\, r + O(p^{1/2+\epsilon}).$$

In particular if $er \geq p^c$, $c > {}^3/_2$, then

$$N(e, r) \sim \frac{\phi(e)}{p - 1}\, r \qquad (p \to \infty).$$

3. In the following theorems α_i, δ_i are fixed non-zero numbers of $GF(p^n)$ and $e_i \mid p^n - 1$.

THEOREM 11. *The number of solutions $\beta_1, \ldots, \beta_r$ of*

$$\alpha = \alpha_1\beta_1 + \ldots + \alpha_r\beta_r, \tag{3.1}$$

where β_i belongs to the exponent e_i is given by

$$N_r(\alpha) = \frac{\theta(e_1) \ldots \phi(e_r)}{p^n - 1} + O(p^{n(1/2+\epsilon)}\, \phi(e_3) \ldots \phi(e_r))$$

for $r \geq 3$ and arbitrary α,

$$N_2(\alpha) = \frac{\phi(e_1)\phi(e_2)}{p^n - 1} + O(p^{n(1/2+\epsilon)}) \qquad (\alpha \neq 0),$$

as $p^n \to \infty$.

In particular for $e_i = p^n - 1$, we have

THEOREM 11′. *The number of solutions of (3.1) in primitive roots is given by*

$$\frac{\phi^r(p^n - 1)}{p^n - 1} + O(p^{n(r - 3/2 + \epsilon)}),$$

for $r \geq 3$ and all α and for $r = 2$, $\alpha \neq 0$.

Let $k_1, \ldots, k_s \geq 2$, $k_i \mid p^n - 1$; let $N_{r,s}(\alpha)$ denote the number of solutions $\beta_1, \ldots, \beta_r, \xi_1, \ldots, \xi_s$ of

$$\alpha = \alpha_1\beta_1 + \ldots + \alpha_r\beta_r + \delta_1\xi^{k_1} + \ldots + \delta_s\xi_s^{k_s},$$

where β_i belongs to the exponent e_i and ξ_i is arbitrary. We now state (compare ref. 3)

THEOREM 12. *For $r \geq 1$, $s \geq 1$, $k_i = O(p^{na})$, $a < 1/2$, we have*

$$N_{r,s}(\alpha) = \phi(e_1) \ldots \phi(e_r)p^{n(s-1)} + O(p^{n(s - 1/2 + a + \epsilon)}\phi(e_2) \ldots \phi(e_r)).$$

The result is valid for all α (except $\alpha = 0$ when $r = s = 1$).

Finally we take $r = 1$, s even, $k_i = 2$, $p > 2$ and prove

THEOREM 13. *The number of solutions $N_{1,2s}(\alpha)$ of*

$$\alpha = \gamma\beta + \delta_1\xi_1^2 + \ldots + \delta_{2s}\xi_{2s}^2 \qquad (\gamma\delta_i \neq 0)$$

is determined by

$$N_{1,2s}(0) = p^{n(2s-1)} + (p^{ns} - p^{n(s-1)})\psi(\zeta),$$

and for $\alpha \neq 0$

$$N_{1,2s}(\alpha) = p^{2ns}(1 - p^{-n}) + p^{ns}\psi(\zeta)\left\{p^{-n} - \frac{\phi(e)}{p^n - 1} + \omega(\alpha)\right\},$$

where $\psi(\zeta) = +1$ or -1 according as $\zeta = (-1)^s\delta_1 \ldots \delta_{2s}$ is or is not a square of $GF(p^n)$, and $\omega(\alpha)$ has the same meaning as in Lemma 2.

For the case of an odd number of squares the method used in the proof of Theorem 13 does not seem to lead to a simple explicit result.

[1] Davenport, H., "On Primitive Roots in a Finite Field," *Quart. J. Math. (Oxford)*, **8**, 308–317 (1937).

[2] Davenport, H., and Hasse, H., "Die Nullstellen der Kongruenzzetafunktionen in gewissen zyklischen Fällen," *J. reine angewandte Math.*, **172**, 151–182 (1935).

[3] Hua, L. K., and Vandiver, H. S., "Characters Over Certain Types of Rings with Applications to the Theory of Equations in a Finite Field," these PROCEEDINGS, **35**, 94–99 (1945).

[4] Landau, E., *Vorlesungen über Zahlentheorie*, Leipzig, 1927, Vol. 2.

[5] Ore, O., "Contributions to the Theory of Finite Fields," *Trans. Am. Math. Soc.*, **36**, 243–274 (1934).

[6] Weil, A., "On the Riemann Hypothesis in Function Fields," these PROCEEDINGS, **27**, 345–347 (1941).

ON ANALYTIC SURFACES WITH TWO INDEPENDENT MEROMORPHIC FUNCTIONS

BY WEI-LIANG CHOW AND KUNIHIKO KODAIRA

JOHNS HOPKINS UNIVERSITY AND INSTITUTE FOR ADVANCED STUDY

Communicated by Oscar Zariski, January 24, 1952

Let V be a compact complex analytic variety of complex dimension r, defined *in abstracto* without any embedding space.[1] According to a recent result of Chow,[2] the field $\mathfrak{F}(V)$ of all meromorphic functions on V is a finite algebraic function field whose degree of transcendency s is at most equal to r. There exists therefore an algebraic variety V^* of dimension s in a projective space, uniquely determined up to a birational transformation, and a meromorphic transformation Φ of V onto V^*, such that every meromorphic function on V is the image of a meromorphic (and hence rational) function on V^* under the inverse transformation Φ^{-1}. Thus the meromorphic transformation Φ induces an isomorphism between the two fields $\mathfrak{F}(V)$ and $\mathfrak{F}(V^*)$, so that we can identify the two and set $\mathfrak{F}(V) = \mathfrak{F}(V^*) = \mathfrak{F}$. The variety V^* will be called the *algebraic* equivalent of V. Now, we shall say that the analytic variety is *algebraic* (in an extended sense), if its algebraic equivalent V^* can be so chosen that the transformation Φ is bi-regular between V and V^*. Since from the abstract point of view two bi-regularly equivalent analytic varieties can be considered as not essentially distinct, it follows that in this case we can identify V with V^*, so that our definition of an algebraic variety is a natural extension of the usual one for a variety in a projective space. It is obvious that an algebraic variety V of dimension r possesses the maximum possible number r of algebraically independent meromorphic functions. The question naturally arises whether the converse is also true: *If a compact analytic variety has as many algebraically independent meromorphic functions as its dimension, is it then necessarily algebraic?* So far as we know, this question has not yet been decided except for some special cases. It is therefore probably not without some interest to present here an answer to this question for the case of a *compact analytic manifold of complex dimension 2 with a Kählerian metric*, which we shall call a Kählerian surface. In the following we shall prove:

THEOREM. *Any Kählerian surface with two algebraically independent meromorphic functions is an algebraic surface without singularities.*

We begin with the observation that the so-called quadratic Cremona transformation in algebraic geometry can be applied also to an analytic surface V. In fact, it can be described in terms of local coordinates as follows: let p be a point on V and (x_1, x_2) be a system of local coordinates with the center p. We consider the direct product $U_p \times S$ of a neighbor-

hood U_p of p and a projective line S and, by means of the homogeneous coordinates (t_0, t_1) on S, construct the subvariety

$$W_p = \{(x_1, x_2;\ t_0, t_1) \mid t_0x_2 - t_1x_1 = 0\} \tag{1}$$

of $U_p \times S$. Obviously W_p is an analytic surface without singularities and contains the projective line $p \times S$ which will be denoted by the same symbol S: $S = p \times S \subset W_p$. The mapping

$$P_p' : (x_1, x_2;\ t_0, t_1) \to (x_1, x_2)$$

is a regular mapping of W_p onto U_p which maps S onto p and $W_p - S$ bi-regularly onto $U_p - p$. Hence, replacing U_p by W_p, we get from V a new analytic surface $V_1 = (V - U_p) \cup W_p$ having the following properties: *There exists a regular mapping P_p of V_1 onto V such that $S = P_p^{-1}(p)$ is a projective line and that P_p is bi-regular between $V_1 - S$ and $V - p$ (P_p is* a "natural" extension of P_p'). The inverse mapping $Q_p = P_p^{-1}$ of P_p will be called the *quadratic transformation with the center p*. It is easy see that the *quadratic transform* $V_1 = Q_p(V)$ of V is determined uniquely by p (and V) and is independent of the choice of the local coordinates (x_1, x_2) appearing in the above construction. Every meromorphic function $F = F(x)$ on V induces a meromorphic function $F(P_p)$ on $Q_p(V)$. Conversely every meromorphic function F_1 on $Q_p(V)$ induces a meromorphic function F on V such that $F(P_p) = F_1$. In fact, it is obvious that F_1 induces on $V - p$ a meromorphic function F such that $F(P_p) = F_1$. To prove that F is meromorphic also at p, consider the meromorphic function $x_1^{-m}F_1$ in a neighborhood of S, m being the multiplicity of the component S in the divisor of F_1. Then, on the curve S, $x_1^{-m}F_1$ equals a rational function $R(t)$ in $t = t_1/t_0$, so that $R(x_2/x_1)^{-1}x_1^{-m}F_1$ is constant on S. This shows that $R(x_2/x_1)^{-1}x_1^{-m}F_1$ induces a holomorphic function $H(x_1, x_2)$ in a neighborhood of p, proving that $F = x_1^m R(x_2/x_1)H(x_1, x_2)$ is meromorphic at p. The mapping $F \to F(P_p)$ gives therefore an isomorphism between $\mathfrak{F}(V)$ and $\mathfrak{F}(Q_p(V))$. Thus *the quadratic transformation Q_p does not affect the field of all meromorphic functions on V: $\mathfrak{F}(Q_p(V)) = \mathfrak{F}(V)$.*

LEMMA 1. *If V is algebraic, then $Q_p(V)$ is also algebraic.*

Proof is well known.[3]

LEMMA 2. *If V is a Kählerian surface, then $Q_p(V)$ is also a Kählerian surface.*

Proof: Using the system of local coordinates (x_1, x_2) with the center p, the Kählerian metric on V can be expressed in a neighborhood U_p of p as

$$ds^2 = \sum_{j,k=1}^{2} (\partial^2 K/\partial x_j \partial \bar{x}_k) dx_j\, d\bar{x}_k,$$

K being a function of class C^∞ defined in U_p. Assuming that U_p is the unit sphere $\{(x_1, x_2) \mid |x_1|^2 + |x_2|^2 < 1\}$ and choosing a function $\psi(\rho)$ of

class C^∞ such that

$$\psi(\rho) = \begin{cases} 1, & \text{for } 0 \leqq \rho \leqq 1/3, \\ 0, & \text{for } \rho \geqq 2/3, \end{cases}$$

we construct a new function

$$\tilde{K} = K + \epsilon \cdot \psi(|x_1|^2 + |x_2|^2) \cdot \log (|x_1|^2 + |x_2|^2),$$

ϵ being a sufficiently small positive constant, and put

$$d\tilde{S}^2 = \sum_{j,k=1}^{2} (\partial^2 \tilde{K} / \partial x_j \partial \bar{x}_k) dx_j \, d\bar{x}_k, \qquad (\text{for } |x_1|^2 + |x_2|^2 > 0).$$

Then, considering (x_1, x_2) as the coordinates of the point $Q_p(x_1, x_2)$ on $W_p = Q_p(U_p)$, we infer readily that $d\tilde{S}^2$ is a Kählerian metric in $W_p - S$. Putting $t = t_1/t_0 = x_2/x_1$, we get

$$ds^2 = ds^2 + \epsilon |dt|^2/(1 + |t|^2)^2, \qquad \text{for } 0 < |x_1|^2 + |x_2|^2 < 1/3,$$

while $\epsilon |dt|^2/(1 + |t|^2)^2$ is a Kählerian metric on S; thus, for $0 < |x_1|^2 + |x_2|^2 < 1/3$, $d\tilde{S}^2$ coincides with the "natural" metric in the Cartesian product $U_p \times S$. This shows that $d\tilde{S}^2$ can be extended to a Kählerian metric in the whole of W_p. On the other hand, $d\tilde{S}^2$ coincides with ds^2 for $|x_1|^2 + |x_2|^2 \geqq 2/3$. Hence $d\tilde{S}^2$ can be extended to a Kählerian metric in the whole of $Q_p(V)$ simply by setting $d\tilde{S}^2 = ds^2$ in $Q_p(V) - W_p$. Thus $Q_p(V)$ is a Kählerian surface.

LEMMA 3. *If $Q_p(V)$ is an algebraic surface, then V is also an algebraic surface.*

Proof: It is sufficient to show that $S = Q_p(p)$ is an *exceptional curve of the first kind.*[4] In fact, if this is the case, there exists a regular mapping P^* of $Q_p(V)$ onto an algebraic surface V^* without singularities such that $P^*(S)$ is a point p^* on V^* and that P^* is bi-regular between $Q_p(V) - S$ and $V^* - p^*$. Then P^*Q_p is a bi-regular mapping of $V - p$ onto $V^* - p^*$ and therefore it must be a bi-regular mapping of V onto V^*. Hence V is algebraic.

Now, in order to prove that $S = Q_p(p)$ is an exceptional curve of the first kind, it is sufficient to show that the intersection number $I(S, S)$ equals -1.[5] Let (x_1, x_2) be the local coordinates on V with the center p. Then x_2 can be considered as a holomorphic function defined in the neighborhood W_p of $S = Q_p(p)$ whose divisor (x_2) consists of S and a curve C such that $I(S, C) = 1$. Now, choose a meromorphic function F on $Q_p(V)$ having S as a polar curve of multiplicity 1 and put $D = (F) + S$, (F) being the divisor of F. Then, since x_2F induces a meromorphic function on S with the divisor $(x_2F) = DS + CS$, we have

$$I(D, S) + I(C, S) = \text{degree of } (x_2F) = 0$$

and therefore

$$I(S, S) = I(D, S) = -I(C, S) = -1,$$

q. e. d.

Now, let V be an arbitrary Kählerian surface such that its algebraic equivalent V^* has the dimension 2; as was mentioned above, we may assume that V and V^* have the same meromorphic functions field $\mathfrak{F} = \mathfrak{F}(V) = \mathfrak{F}(V^*)$. Let $(X_0, X_1, \ldots, X_n)$ be the homogeneous coordinates of a generic point X of V^*. Then, by means of meromorphic functions $F_k(x) = X_k/X_0$ of a "variable" point x on V, the meromorphic mapping Φ of V onto V^* can be represented as

$$\Phi: x \to \Phi(x) = (1, F_1(x), \ldots, F_n(x)).$$

Denote by C_λ the inverse image $\Phi^{-1}(C_\lambda^*)$ of the hyperplane section C_λ^* of V^* cut out by any hyperplane $\lambda_0 X_0 + \lambda_1 X_1 + \ldots + \lambda_n X_n = 0$. Then the set $|C_\lambda|$ of all such curves C_λ constitutes a *linear system* on V having no fixed component and the meromorphic mapping $x \to \Phi(x)$ is *regular* except for a finite number of *base points* of this linear system $|C_\lambda|$. By the *multiplicity* of a base point p of $|C_\lambda|$ we shall mean the integer

$$m(p) = \min_{\lambda, \mu} I_p(C_\lambda, C_\mu),$$

where $I_p(C_\lambda, C_\mu)$ denotes the intersection multiplicity of C_λ and C_μ *at* p. Obviously $m(p)$ is always not smaller than 1 (or $m(p)$ equals zero if and only if p is *not* a base point). Now we shall show that these base points can be eliminated by means of quadratic transformations (if any base point ever exists). Take a base point p of $|C_\lambda|$ and construct the quadratic transform $V' = Q_p(V)$. The mapping Φ induces a mapping Φ': $(x') \to \Phi'(x') = \Phi Q_p^{-1}(x')$ of V' onto V^* which corresponds to the linear system $|C_\lambda'|$ of curves $C_\lambda' = Q_p(C_\lambda)$ on V'. This new system $|C_\lambda'|$ might have a base point p' on $S = Q_p(p)$, but the *multiplicity* $m(p')$ *of* p' *is strictly smaller than* $m(p)$, since we have

$$I_{p'}(C_\lambda', C_\mu') = I_p(C_\lambda, C_\mu) - m_\lambda \cdot m_\mu,^6$$

m_λ and m_μ being the multiplicities of the curves C_λ and C_μ at p, respectively. This shows clearly that, by applying a finite number of suitable quadratic transformations Q_p, $Q_{p'}$, ... to V successively, we get an analytic surface

$$\tilde{V} = \ldots Q_{p''} Q_{p'} Q_p(V) \qquad (2)$$

such that the mapping $\tilde{\Phi} = \Phi Q_p^{-1} Q_{p'}^{-1} Q_{p''}^{-1} \ldots$ of $\tilde{V}$ onto *its algebraic equivalent* V^* is a regular mapping. By virtue of Lemma 2, $\tilde{V}$ *is also a Kählerian surface.*

The inverse mapping $\tilde{\Phi}^{-1}$ has possibly a finite number of fundamental

points and branch curves and, except for these "singular" points, $\tilde{\Phi}^{-1}$ is locally bi-regular. Hence, for any point p^* on V^* which is not "singular" with respect to $\tilde{\Phi}^{-1}$, the set $\tilde{\Phi}^{-1}(p^*)$ consists of a finite number d of points on $\tilde{V}$ (the number d is independent of p^*). Now we shall show that $d = 1$. Let E_h^* be the section of V^* cut out by a "general" hypersurface of order h containing no fundamental point of $\tilde{\Phi}^{-1}$ and put $E_h = \tilde{\Phi}^{-1}(E_h^*)$. Then, since V^* is the algebraic equivalent of $\tilde{V}$, the complete linear systems $|E_h|$ on $\tilde{V}$ and $|E_h^*|$ on V^* have the same dimension: $\dim|E_h| = \dim|E_h^*|$. For sufficiently large h the system $|E_h^*|$ is *regular* and therefore it follows from *Riemann-Roch's theorem*[7] that

$$\dim|E_h^*| = {}^1/_2 I(E_h^*, E_h^*) - {}^1/_2 I(K^*, E_h^*) + p_a^*,$$

where K^* is the canonical divisor on V^* and p_a^* is the arithmetic genus of V^*, while, using *Riemann-Roch's theorem on Kählerian surfaces* proved recently by Kodaira,[8] we get

$$\dim|E_h| \geqq {}^1/_2 I(E_h, E_h) - {}^1/_2 I(K, E_h) + p_a - i,$$

where K is the canonical divisor on $\tilde{V}$, p_a is the arithmetic genus of $\tilde{V}$ and $i = \dim|K - E_h| + 1$. Putting $\nu = I(E_1^*, E_1^*)$, $k^* = I(K^*, E_1)$, $k = I(K, E_1)$, we get

$$\dim|E_h^*| = {}^1/_2 \nu \cdot h^2 - {}^1/_2 k^* \cdot h + p_a^*$$

and, since $I(E_h, E_h) = d \cdot I(E_h^*, E_h^*)$,

$$\dim|E_h| \geqq {}^1/_2 d \cdot \nu \cdot h^2 - {}^1/_2 k \cdot h + p_a - \dim|K| - 1.$$

Combined with $\dim|E_h^*| = \dim|E_h|$, the above two formulae show clearly that *d must be equal to* 1. It follows from $d = 1$ that $\tilde{\Phi}^{-1}$ has no branch curve; thus *$\tilde{\Phi}^{-1}$ is regular and single-valued except for a finite number of fundamental points. $\tilde{\Phi}$ is therefore bi-regular on $\tilde{V}$ except for a finite number of fundamental curves.*

Now we shall show that these fundamental curves can be eliminated by applying quadratic transformations to V^*. Let p^* be a fundamental point on V^*. Then $L = \tilde{\Phi}^{-1}(p^*)$ is a (possibly reducible) fundamental curve on $\tilde{V}$. Let $L = L_1 + L_2 + \ldots + L_n$ be the decomposition of L into a sum of irreducible components L_j. Now, let F_1, F_2 be two meromorphic functions in $\mathfrak{F} = \mathfrak{F}(\tilde{V}) = \mathfrak{F}(V^*)$ such that (F_1, F_2) constitutes a system of local coordinates on V^* with the center p^*. Then the curves B_λ of zeros of the meromorphic functions $\lambda_1 F_1 + \lambda_2 F_2$ on $\tilde{V}$ constitute a linear pencil $|B_\lambda|$ on $\tilde{V}$. Apparently the part of the fixed component B_0 of $|B_\lambda|$ lying on L is written as $B_0 = \sum k_j L_j$ so that we get

$$B_\lambda = \sum k_j L_j + B_{\lambda 1}.$$

The same argument as in the case of fundamental curves on algebraic

surfaces[9] shows that the curve $B_{\lambda 1}$ meets with one irreducible component, say L_1, of L in one "variable" point with intersection multiplicity 1 (i.e., $I(B_{\lambda 1}, L_1) = 1$) but does not meet other components L_2, L_3, ... of L. This implies that, at each point p on L, the ratio F_2/F_1 can be expressed as

$$F_2/F_1 = H_{2p}/H_{1p}$$

by means of two holomorphic functions H_{1p}, H_{2p} in a neighborhood of p not vanishing simultaneously and that H_{2p}/H_{1p} is a meromorphic function on L_1 whose value $\lambda = H_{2p}/H_{1p}$ corresponds one-to-one to the intersection point $B_{\lambda 1}L_1$ on L_1 while H_{2p}/H_{1p} is constant on other components L_2, L_3, ... Now, consider the mapping $Q_{p*}\tilde{\Phi}$ of $\tilde{V}$ onto the quadratic transform $Q_{p*}(V^*)$ which is also an algebraic surface, as Lemma 1 shows. By means of the local coordinates $(x_1^*, x_2^*; t_0^*, t_1^*)$ on $Q_{p*}(V^*)$ introduced in (1), this mapping $Q_{p*}\tilde{\Phi}$ can be represented in a neighborhood of L as

$$(x_1^*, x_2^*; t_0^*, t_1^*) = (F_1, F_2; H_{p1}, H_{p1});$$

hence $Q_{p*}\tilde{\Phi}$ is a regular mapping. $Q_{p*}\tilde{\Phi}$ maps L_1 onto $S^* = Q_{p*}(p^*)$ one-to-one. Consequently L_1 is not a fundamental curve of $Q_{p*}\tilde{\Phi}$, whereas L_2, L_3, ... are still fundamental curves of $Q_{p*}\tilde{\Phi}$; again it is obvious that no new fundamental curve is created by Q_{p*}. Thus the quadratic transformation Q_{p*} eliminates one irreducible component L_1 of the fundamental curve L. This shows that, by applying a finite number of suitable quadratic transformations Q_{p*}, $Q_{p*'}$, ... to V^*, we get an algebraic surface

$$V_1^* = \ldots Q_{p*'}Q_{p*}(V^*)$$

such that the mapping $\tilde{\Phi}_1 = \ldots Q_{p*'}Q_{p*}\tilde{\Phi}$ of $\tilde{V}$ onto V_1^* is bi-regular. This proves that $\tilde{V}$ is an algebraic surface. Now $\tilde{V}$ is obtained from V by a finite number of quadratic transformations, as (2) shows. Hence, using Lemma 3, we conclude that *the given Kählerian surface V is an algebraic surface*, q. e. d.

[1] For the definition of a complex analytic variety *in abstracto*, see Chow, W. L., "On Complex Analytic Varieties," to appear in the *Am. J. Math.* For the purpose of this paper we can restrict ourselves to complex analytic manifolds in the usual sense.

[2] See Chow, W. L., *loc. cit.*, Theorem 4.6. We mention here that a closed analytic correspondence Φ of an analytic variety V into a compact analytic variety U is a meromorphic transformation, if it is single-valued in general and if any meromorphic function (in the large) on U whose polar variety does not contain $\Phi(V)$ is carried by Φ^{-1} into a meromorphic function on V. In case V is an algebraic variety embedded in a projective space, this is equivalent to saying that Φ can be represented by setting the inhomogeneous coordinates of a point in V equal to a system of meromorphic functions on U.

[3] See, for example, van der Waerden, B. L., *Infinitely Near Points*, Koninklijke Nederlandse Akademie van Wetenschappen, Vol. LIII, 401–410, § 3 (1950).

[4] Cf. Zariski, *Algebraic Surfaces*, Berlin, pp. 36–41 (1935).

[5] Using Riemann-Roch's theorem it can be shown that a rational curve S without singularities is an exceptional curve of the first kind if $I(S, S) = -1$. See Zariski,

loc. cit., pp. 71–72. Strictly speaking this criterion holds only for an algebraic surface in the usual sense, i.e., embedded in a projective space. However, it can be easily extended to an algebraic V in the extended sense by observing that if V is mapped by a bi-regular transformation Φ onto an algebraic surface V' in a projective space, then any curve S in V is mapped by Φ bi-regularly onto a compact analytic curve S' in the projective space, which must then necessarily be an algebraic curve. See Chow, W. L., "On Compact Complex Analytic Varieties," *Am. J. Math.*, **71**, 893–914 (1949), Theorem V. It is clear that if S is a rational curve without singularities in V such that $I(S, S) = -1$, then S' is also a rational curve without singularities in V' such that $I(S', S') = -1$. It follows then that S' is an exceptional curve of the first kind in V', i.e., there exists a bi-rational transformation Ψ of V' which has S' as a fundamental curve and is bi-regular everywhere else. Then the "product" transformation $\Psi\Phi$ is a bi-rational transformation of V which has S as a fundamental curve and is bi-regular everywhere else.

[6] See van der Waerden, *loc. cit.*, § 3.

[7] Cf. Zariski, *loc. cit.*, pp. 66–71.

[8] Kodaira, K., "The Theorem of Riemann-Roch on Compact Analytic Surfaces," *Am. J. Math.*, **73**, 813–875 (1951), Theorem 5.3.

[9] See Zariski, *loc. cit.*, pp. 36–41.

COHOMOLOGY GROUPS OF ABELIAN GROUPS AND HOMOTOPY THEORY IV

BY SAMUEL EILENBERG* AND SAUNDERS MACLANE

DEPARTMENTS OF MATHEMATICS, COLUMBIA UNIVERSITY AND THE UNIVERSITY OF CHICAGO

Communicated February 27, 1952

The present note is concerned with extension and classification theorems for continuous mappings.

1. Products.—Let K be a simplicial complex with ordered vertices,[1] $L_1, \ldots, L_r$ subcomplexes of K and $L = L_1 \cup \ldots \cup L_r$. Let $z^n \in Z^n(K; \Pi)$ $(n \geqq 1)$ be an n-cocycle with values in an abelian group Π. If σ_q is a q-simplex of K then z^n determines a cocycle in $Z^n(\Delta_q, \Pi)$ which may be regarded as a q-simplex of the complex $K(\Pi, n)$.[2] There results a simplicial map $T(z^n)\colon K \to K(\Pi, n)$. If z^n is zero on L_i then $T(z^n)$ maps L_i into the subcomplex $K(0, n)$ of $K(\Pi, n)$. Cohomologous cocycles yield chain-homotopic maps.

Let $x_i \in H^n(K, L_i;\ \Pi)$, $i = 1, \ldots, r$ and let $y \in H^q(\Pi, n;\ G)$. We define the product

$$[x_1, \ldots, x_r;\ y] \in H^q(K, L;\ G)$$

as follows: Let $z_i \in x_i$ be cocycles. Consider the chain transformation $s\colon (K, L) \to (K(\Pi, n), K(0, n))$ defined by

$$s = \sum(-1)^{r-k}T(z_{i_1} + \ldots + z_{i_k}), \qquad 1 \leqq i_1 < \ldots < i_k = r;$$

then

$$[x_1, \ldots, x_r;\ y] = s^*y.$$

This product is additive in y independent of the order of $x_1, \ldots, x_r$, and is natural with respect to simplicial maps of K, homomorphisms of Π and homomorphisms of G. Further properties are:

$$[x_1, \ldots, x_{r-1}, x_r + x_r';\ y] = [x_1, \ldots, x_{r-1}, x_r;\ y] + [x_1, \ldots, x_{r-1}, x_r';\ y] + [x_1, \ldots, x_{r-1}, x_r, x_r';\ y], \quad (1)$$

$$[x_1, \ldots, x_r;\ y] = 0, \quad \text{if} \quad q < rn, \quad (2)$$

$$[x_1, \ldots, x_r;\ y] = x_1 \cup \ldots \cup x_r, \text{ if } q = rn. \quad (3)$$

In (3) the cup product is to be taken relative to the pairing $\Pi \otimes \ldots \otimes \Pi \to G$ which to each system $\alpha_1, \ldots, \alpha_r \in \Pi$ assigns the value of y on the cocycle $\alpha_1^* \ldots {}^*\alpha_r$, where $*$ denotes the product of chains in the complex $K(\Pi, n)$.[3] For $r = 1$ we have the following additional properties. Let $b^n \in H^n(\Pi, n; \Pi)$ be the basic cohomology class defined by $b^n(\alpha) = \alpha$, $\alpha \in \Pi$. Then

$$y = [b^n;\ y]. \quad (4)$$

Let $x \in H^n(L; \Pi)$, $\delta x \in H^{n+1}(K, L; \Pi)$, $y \in H^{q+1}(\Pi, n + 1; G)$. Then

$$[\delta x;\ y] = \delta[x;\ Sy], \quad (5)$$

where $S: H^{q+1}(\Pi, n + 1; G) \to H^q(\Pi, n; G)$ is the suspension homomorphism.

2. *Depressed Products.*—Let $K \times I$ be the cartesian product of K with the closed unit interval, subdivided simplicially in the usual manner. For a subcomplex L of K we define

$$\tilde{L} = L \times I \cup K \times 0 \cup K \times 1, \quad L^1 = L \times I \cup K \times 1.$$

We define maps

$$p: (K \times I, L \times I) \to (K, L), \quad l: (K, L) \to (\tilde{L}, L^1)$$

by $p(u, t) = u$, $l(u) = (u, 0)$. Both p^* and l^* are isomorphisms for cohomology groups. Combining l^{*-1} with

$$\delta: H^q(\tilde{L}, L^1) \to H^{q+1}(K \times I, \tilde{L})$$

we obtain an isomorphism

$$\varphi: H^q(K, L) \approx H^{q+1}(K \times I, \tilde{L})$$

for any coefficient group.

Now, let $x_i \in H^n(K, L; \Pi)$ $i = 1, \ldots, r - 1$, $x_r \in H^{n-1}(K, L_r; \Pi)$, $y \in$

$H^q(\Pi, n; G)$. We define the *depressed* product

$$\langle x_1, \ldots, x_{r-1}, x_r; y\rangle = \varphi^{-1}[p_1{}^*x_1, \ldots p_{r-1}^*x_{r-1}, \varphi x_r; y]$$

which is an element of $H^{q-1}(K, L; G)$. As in (2) and (3) we have

$$\langle x_1, \ldots, x_r; y\rangle = 0 \quad \text{if} \quad q < rn \tag{6}$$

$$\langle x_1, \ldots, x_r; y\rangle = x_1 \cup \ldots \cup x_r \quad \text{if } q = rn.^5 \tag{7}$$

Further

$$\langle x; y\rangle = [x; Sy]. \tag{8}$$

The products above may be applied to the case when $K = S(X)$ is the singular complex of a topological space, thus carrying over the products to topological spaces. The invariance of the products in a simplicial complex may be established similarly.

3. *Obstruction Theorems.*—Let Y be a pathwise connected topological space with

$$\pi_i(Y) = 0 \text{ for } 1 \leqq i < n \quad \text{and} \quad n < i < q$$

where $1 < n < q$. We consider the natual homomorphisms[6]

$$\kappa\colon H^k(\pi_n, n; G) \to H^k(Y; G), \qquad \pi_n = \pi_n(Y),$$

and introduce the basic cohomology class

$$\iota^n = \kappa^* b^n \in H^n(Y; \pi_n).$$

The space Y has an obstruction

$$k^{q+1} = k^{q+1}(Y) \in H^{q+1}(\pi_n, n; \pi_q).$$

Let $f\colon K^n \cup L \to Y$ be a map extendable to $K^{n+1} \cup L \to Y$. Then f admits an extension $f'\colon K^q \cup L \to Y$. The obstruction $c^{q+1}(f') \in H^{q+1}(K, L; \pi_q)$ is then independent of f' and is denoted by $z^{q+1}(f)$. This is the *secondary obstruction*[7] of f. The homomorphisms $f'^*\colon H^n(Y; G) \to H^n(K; G)$ are also independent of f' are denoted by f^*. If $g\colon K^n \cup L \to G$ is another map with extension $g'\colon K^q \cup L \to Y$ such that $f|L = g|L$, then the difference homomorphism

$$(f' - g')^*\colon H^n(Y; G) \to H^n(K, L; G)$$

is independent of f' and g' and will be written as $(f - g)^*$.

THEOREM I. *Let $f, g\colon K^n \cup L \to Y$ be maps extendable to $K^{n+1} \cup L \to Y$ and such that $f|L = g|L$. Then*

$$z^{q+1}(f) - z^{q+1}(g) = [(f - g)^*\iota^r; k^{q+1}] + [g^*\iota^n, (f - g)^*\iota^n; k^{q+1}]. \tag{9}$$

THEOREM II. *Let $f_0, f_1\colon K^q \cup L \to Y$ be two maps with $f_0|K^{q-1} \cup L =$*

$f_1 | K^{q-1} \cup L$ *and with the difference cohomology class*[7] $d^q(f_0, f_1) \in H^q(K, L; \pi_q)$. *Then* f_0 *and* f_1 *are homotopic rel.* L *if and only if*

$$d^q(f_0, f_1) + [e^{n-1}; Sk^{q+1}] + \langle f_0{}^*\iota^n, e^{n-1}; k^{q+1} \rangle = 0 \qquad (10)$$

for some $e^{n-1} \in H^{n-1}(K, L; \pi_n)$.

The assumption that $\pi_i(Y) = 0$ for $i < n$ may be dropped provided one assumes that the various maps carry K^{n-1} into the base point y_0 of Y.

The case $n = 1$, $\pi_1(Y)$ abelian, may be included provided the cohomology groups are taken with local coefficients. In the nonabelian case further changes are necessary and only partial results are obtained. These include earlier known results.[8]

4. The Case $q = n + 1$, $n > 1$.—A better understanding of this case depends upon a closer examination of the products

$$[x_1; y], \qquad (11)$$

$$[x_1, x_2; y], \qquad (12)$$

$$[x_3; Sy], \qquad (13)$$

$$\langle x_1, x_2; y \rangle, \qquad (14)$$

where

$$x_1 \in H^n(K, L_2; \Pi), x_2 \in H^n(K, L_2; \Pi),$$

$$x_3 \in H^{n-1}(K, L_3; \Pi), y \in H^{n+2}(\Pi, n; G).$$

This is possible because of our good knowledge of the group $H^{n+2}(\Pi, n; G)$. Indeed the elements of this group are in a 1-1-correspondence with functions t: $\Pi \to G$ (called *traces*) satisfying

$$t(-\alpha) - t(\alpha) = 0, \qquad (15)$$

$$t(\alpha + \beta + \gamma) - t(\beta + \gamma) - t(\alpha + \gamma) - t(\alpha + \beta) + t(\alpha) + t(\beta) + t(\gamma) = 0, \text{ if } n = 2, \qquad (16)$$

$$t(\alpha + \beta) - t(\alpha) - t(\beta) = 0, \qquad \text{if } n > 2. \qquad (17)$$

The products (12) and (14) are easiest to interpret. They are zero for $n > 2$ and are, respectively, the cup products $x_1 \cup x_2$ and $x_1 \cup x_3$ if $n = 2$. The cup product is taken relative to the pairing $t(\alpha + \beta) - t(\alpha) - t(\beta)$ where t is the trace of y.

If $n > 2$, then there exist a bilinear function d: $\Pi \otimes \Pi \to G$ such that $d(\alpha, \alpha) = t(\alpha)$. Relative to such a *bilinear extension* d of t one may consider the Steenrod products.[7] It turns out that (11) and (13) reduce to Sq^2x_1 and Sq^2x_3, respectively.

For $n = 2$, we consider an exact sequence

$$0 \to \Pi'' \xrightarrow{\psi} \Pi' \xrightarrow{\varphi} \Pi \to 0$$

and define a *generalized bilinear extension* of t to be a bilinear function d: $\Pi' \otimes \Pi' \to G$ such that

$$d(\alpha, \alpha) = t(\varphi\alpha) \qquad \text{for } \alpha \in \Pi'$$

$$d(\psi(\beta), \psi(\gamma)) = 0 \qquad \text{for } \beta, \gamma \in \Pi''.$$

Such a d always exists if Π' is free and Π is finitely generated. Let

$$u \in C^2(K, L_1, \Pi'), \qquad v \in C^1(K, L_3, \Pi')$$

be cochains such that φu and φv are cocycles in the cohomology classes x_1 and x_3, respectively. Then the expressions

$$u \cup u - u \cup_1 \delta u, \quad v \cup \delta v$$

where $\cup$ and $\cup_1$ are taken relative to d, are cocycles in the class (11) and (13), respectively.

In the particular case $y = k^{n+2}(Y)$, the trace t of y coincides with J. H. C. Whitehead's function η: $\pi_n \to \pi_{n+1}$ obtained by combining each map $S^n \to Y$ with the map $S^{n+1} \to S^n$ which is the Hopf map for $n = 2$ and its $(n-2)$-fold suspension for $n > 2$.[9]

The above results include those of Steenrod,[7] Whitney,[10] Postnikov[11] and J. H. C. Whitehead.[12]

* Essential portions of the study here summarized were done during the tenure of a John Simon Guggenheim Fellowship by one of the authors.

[1] The appropriate class of complexes to work with in §§1, 2 and 4 are the "semi-simplicial" complexes of Eilenberg and Zilber, *Ann. Math.*, **51,** 499–513 (1950).

[2] Eilenberg, S., and MacLane, S., PROC NATL. ACAD. SCI., **36,** 443–447 and 657–663 (1950); **37,** 307–310 (1951); subsequently referred to as Notes I, II and III.

[3] Note III, p. 308.

[4] Note III, p. 307, or Note I, pp. 444–445.

[5] The process of depressing a product may be iterated and may be applied to other products besides those of §1. In particular, the cup product depressed is again the cup product.

[6] Eilenberg, S., and MacLane, S., *Ann. Math.*, **51,** 514–533 (1950).

[7] Steenrod, N. E., *Ibid.*, **48** 290–320 (1947).

[8] Robbins, H., *Trans. Am. Math. Soc.*, **49,** 308–324 (1941). Olum, P. (to appear).

[9] Cf. G. W. Whitehead, PROC. NATL. ACAD. SCI., **34,** 207–211 (1948), Theorem 5.

[10] *Ann. Math.*, **50,** 270–284 (1949).

[11] *C. R. (Doklady) Acad. Sci. URSS N. S.*, **67,** 427–430 (1949).

[12] *Ann. Math.*, **54,** 68–84 (1951).

THE APPARITION PROBLEM FOR EQUIANHARMONIC DIVISIBILITY SEQUENCES

BY L. K. DURST

THE RICE INSTITUTE, HOUSTON

Communicated by E. T. Bell, February 27, 1952

The apparition problem consists in finding an arithmetical relation between a rational prime p and its rank of apparition in an elliptic divisibility sequence.[1] In this report the elliptic divisibility sequences parameterized by equianharmonic functions are investigated, and a partial answer to the apparition problem for such sequences is given (Theorem 2).

If $\omega_2/\omega_1 = \rho = {}^1/_2(-1 + \sqrt{-3})$, the Weierstrass functions for the parallelogram $0, 2\omega_1, 2\omega_2, 2\omega_1 + 2\omega_2$ are called equianharmonic functions and have the following expansions:

$$\begin{aligned}
\sigma(u) &= u \prod_{\mu \text{ in } E}^{\mu \neq 0} \left\{\left(1 - \frac{u}{2\mu\omega_1}\right) \exp\left(\frac{u}{2\mu\omega_1} + \frac{1}{2}\frac{u^2}{(2\mu\omega_1)^2}\right)\right\} \\
\zeta(u) &= \frac{1}{u} + \sum_{\mu \text{ in } E}^{\mu \neq 0} \left\{\frac{1}{u - 2\mu\omega_1} + \frac{1}{2\mu\omega_1} + \frac{u}{(2\mu\omega_1)^2}\right\} \\
\wp(u) &= \frac{1}{u^2} + \sum_{\mu \text{ in } E}^{\mu \neq 0} \left\{\frac{1}{(u - 2\mu\omega_1)^2} - \frac{1}{(2\mu\omega_1)^2}\right\} \\
\wp'(u) &= -2 \sum_{\mu \text{ in } E} \frac{1}{(u - 2\mu\omega_1)^3}
\end{aligned} \tag{1}$$

where the index μ runs through the ring E of the quadratic integers $\mu = a + b\rho$, a and b rational integers. The norm in E is given by $N\mu = N(a + b\rho) = a^2 - ab + b^2$, and the units in E are $\epsilon = \pm 1, \pm\rho, \pm\rho^2 = \mp(1 + \rho)$. Equations (1) imply

$$\begin{aligned}
\sigma(\epsilon u, \omega_1, \rho\omega_1) &= \epsilon\, \sigma(u, \omega_1, \rho\omega_1) \\
\zeta(\epsilon u, \omega_1, \rho\omega_1) &= \epsilon^{-1}\zeta(u, \omega_1, \rho\omega_1) \\
\wp(\epsilon u, \omega_1, \rho\omega_1) &= \epsilon^{-2}\wp(u, \omega_1, \rho\omega_1) \\
\wp'(\epsilon u, \omega_1, \rho\omega_1) &= \epsilon^{-3}\wp'(u, \omega_1, \rho\omega_1)
\end{aligned} \tag{2}$$

where $N\epsilon = 1$. The function $\sigma(u)$ is a doubly periodic function of the third kind:

$$\sigma(u + 2\mu\omega_1) = (-1)^{N\mu}\, e^{2\bar{\mu}\eta_1(u\, +\, \mu\omega_1)}\, \sigma(u), \mu \text{ in } E, \tag{3}$$

$\bar{\mu}$ being the conjugate at μ, and $\eta_1 = \zeta(\omega_1)$. Furthermore (2) implies $g_2 = 0$ and hence

$$\wp'(u)^2 = 4\wp(u)^3 - g_3.$$

Equianharmonic functions afford one of the simpler examples of elliptic functions admitting complex multiplication.[2]

If $\psi_\mu(u) = \sigma(\mu u, \omega_1, \rho\omega_1)\ \sigma(u, \omega_1, \rho\omega_1)^{-N\mu}$, for μ in E, then (1) and (3) imply that $\psi_\mu(u)$ is an elliptic function of u with periods $2\omega_1$ and $2\rho\omega_1$. The three-term sigma formula[3] implies

$$\epsilon^2\psi_{\mu+\nu}(u)\psi_{\mu-\nu}(u) = \psi_{\mu+\epsilon}(u)\psi_{\mu-\epsilon}(u)\psi_\nu^2(u) - \psi_{\nu+\epsilon}(u)\psi_{\nu-\epsilon}(u)\psi_\mu^2(u), \tag{4}$$

for any μ, ν, ϵ in E, provided $N\epsilon = 1$.

1°. Using $\psi_{\epsilon\mu}(u) = \epsilon\psi_\mu(u)$, $N\epsilon = 1$, and the recursion (4), every value of $\psi_\mu(u)$ may be computed from the initial values

$$\psi_0(u) = 0,\ \psi_1(u) = 1,\ \psi_{1-\rho}(u) = (1-\rho)\wp(u),\ \psi_2(u) = -\wp'(u).$$

2°. If $N\mu$ is an odd rational integer,

$$\psi_\mu(u) = P_\mu(z, g_3),$$

and if $N\mu$ is even,

$$\psi_\mu(u) = \wp'(u)P_\mu(z, g_3),$$

where $z = \wp(u)$ and $P_\mu(z, g_3)$ is a polynomial in z over the polynomial ring $E[g_3]$ of degree $\frac{1}{2}(N\mu - 1)$ if $N\mu$ odd, and $\frac{1}{2}(N\mu - 4)$ if $N\mu$ even.

Proofs of 1° and 2° may be given by inductions on $N\mu$.

3°. If $\nu \mid \mu$ in the ring E, then $P_\nu(z, g_3) \mid P_\mu(z, g_3)$ in the ring $E[z, g_3]$.

Let $\mu = \lambda\nu$. Then

$$\psi_\mu(u) = \frac{\sigma(\lambda\nu u)}{\sigma(u)^{N(\lambda\nu)}} = \frac{\sigma(\lambda\nu u)}{\sigma(\nu u)^{N\lambda}}\left\{\frac{\sigma(\nu u)}{\sigma(u)^{N\nu}}\right\}^{N\lambda} = \psi_\lambda(\nu u)\psi_\nu(u)^{N\lambda},$$

and if $N\lambda$ and $N\nu$ are both odd,

$$P_\mu(z, g_3) = P_\lambda(\wp(\nu u), g_3)P_\nu(z, g_3)^{N\lambda}.$$

But

$$\wp(\nu u) = \wp(u) - \psi_{\nu-1}(u)\psi_{\nu+1}(u)\psi_\nu(u)^{-2}, \tag{5}$$

so $P_\lambda(\wp(\nu u), g_3)P_\nu(z, g_3)^{N\lambda-1}$ is a polynomial in $E[z, g_3]$. The details of the proof of 3° are similar if one or both of $N\lambda$, $N\nu$ are even.

If z and g_3 are fixed in the ring I of rational integers, then the correspondence $\mu \to P_\mu(z, g_3)$ is a mapping of E into itself preserving division. Let $\mathfrak{p}$ be a prime ideal of E. An integer λ of E will be called a *zero* of $\mathfrak{p}$ if

$$P_\lambda(z, g_3) \equiv 0 \pmod{\mathfrak{p}}, \qquad z, g_3 \text{ fixed in } I.$$

A zero α of $\mathfrak{p}$ with minimum positive norm will be called a *rank of apparition* of $\mathfrak{p}$. That every $\mathfrak{p}$ of E has a rank of apparition for each z, g_3 of I, may be

shown by an argument similar to the proof of Theorem 5.1 of Ward's Memoir.[1]

Write $P_\mu(z) = P_\mu(z, g_3)$ for g_3 fixed in I. If $M(\delta)$ is the Möbius function of the principal ideal ring E, defined by

$M(\epsilon) = 1$ if $N\epsilon = 1$
$M(\delta) = (-1)^r$ if (δ) is a product of r distinct prime ideals
$M(\delta) = 0$ if (δ) is divisible by the square of a prime ideal,

then

$$Q_\mu(z) = \prod_{(\delta)|(\mu)} P_{\mu/\delta}(z)^{M(\delta)}$$

is in $E[z]$; and by the inversion formula of Dedekind,

$$P_\mu(z) = \prod_{(\delta)|(\mu)} Q_\delta(z),$$

up to a unit factor in E. Since $\sigma(u) = 0$ if and only if $u = 2\nu\omega_1$, ν in E, the roots of $Q_\mu(z) = 0$ are the distinct values of $\wp(2\nu\omega_1/\mu)$, $(\nu, \mu) = 1$.

4°. If α is a rank of apparition of $\mathfrak{p}$, then $Q_\alpha(z)$ and $P_\alpha(z)$ split into linear factors in $E/\mathfrak{p}$.

4° follows from (5).

Let R be the field of rationals and let F_μ be the root field of $Q_\mu(z)$. If $\nu|\mu$, $N\nu > 1$, then $F_\mu \supset F_\nu \supset R(\rho)$, unless $N\nu = 4$. Let $C_n(x) = 0$ be the equation, irreducible over R, satisfied by the primitive nth roots of unity. If $n \neq 3$, $C_n(x)$ is irreducible over $R(\rho)$.

THEOREM 1. *If n is an odd rational integer, then $C_n(x)$ is reducible in F_n.*

For if $\theta = \exp(2\pi i/n)$ and $n = d\,d'$, then

$$\sum_{s=0}^{d-1} \theta^{2rsd'} \wp'(2\omega_1(r + s\rho)/d)^{-1} = 0$$

$$\sum_{s=0}^{d-1} \theta^{2rsd'} \wp(2\omega_1(r + s\rho)/d)\wp'(2\omega_1(r + s\rho)/d)^{-1} = 0 \qquad (6)$$

where $r = 1, 2, \ldots, {}^1/_2(d-1)$.[4] It follows from (5) that multiplication of equations (6) by $\wp'(2\omega_1/n)$ yields $\sum_{d|n}(d-1)$ equations over F_n, of degree at most $n-1$, each of which has θ as a root. Since

$$\phi(n) > n - 1 - \sum_{d|n}(d-1),$$

$C_n(x)$ factors in F_n.

It is now possible to formulate an answer to the apparition problem for all rational primes that do not split in E.

THEOREM 2. *If $p \equiv 2 \pmod 3$ and α is a rank of apparition of p, then*

$$\alpha = 2^e b \qquad \text{or} \qquad \alpha = 2^e b(1 - \rho)$$

where b is 1, 3, or an odd divisor of $p^e - 1$ and c and e are rational integers, $c \geq 0$ and $e < \phi(b)$.

By hypothesis $(p) = \mathfrak{p}$, a prime ideal of E. But $\alpha = \bar{\alpha}\epsilon$ since $\mathfrak{p} = \bar{\mathfrak{p}}$, so that

$$\alpha = a \qquad \text{or} \qquad \alpha = a(1 - \rho),$$

where a is a rational integer. Let $a = 2^c b$, b odd. Since $b \mid \alpha$,

$$F_\alpha \supseteq F_b \supseteq R(\rho), \qquad \text{if } b > 1.$$

Hence by 4° and Theorem 1, $C_b(x)$ is reducible in E/p; and if $b \neq 3$, $p^e \equiv 1 \pmod{b}$, where e is the common degree of the irreducible factors of $C_b(x)$ in E/p.

If $p \equiv 1 \pmod{3}$, then $p = N\mathfrak{p}$, $\mathfrak{p}$ a prime ideal of E. In this case the apparition problem is an open question.

[1] Ward, Morgan, "Memoir on Elliptic Divisibility Sequences," *Am. J. Math.*, **70**, 31–74 (1948).

[2] Ward, Morgan, "Arithmetical Properties of Polynomials Associated with the Lemniscate Elliptic Functions," PROC. NATL. ACAD. SCI., **36**, 359–362 (1950).

[3] For the three-term sigma formula, and all of the formulas of the previous paragraph, cf. Tannery and Molk, *Éléments de la théorie des fonctions elliptiques*, Vol. 2, pp. 234–236.

[4] For these "Abelian Relations," cf. Fricke, *Die elliptischen Funktionen und ihre Anwendungen*, Vol. 2, p. 242.

SOME THEOREMS ON PIECEWISE LINEAR EMBEDDING

BY V. K. A. M. GUGENHEIM

MAGDALEN COLLEGE, OXFORD

Communicated by S. Lefschetz, January 30, 1952

1. We call two (Euclidean) polyhedra equivalent if they have isomorphic simplicial subdivisions; the homeomorphism which maps each simplex of such a subdivision linearly onto its correlate in an isomorphic subdivision is called a piecewise linear homeomorphism onto, abbreviated *PLO*.

A polyhedron is called finite if it has a subdivision consisting of a finite number of simplices; a polyhedron equivalent to a q-simplex is called a q-element, one equivalent to the boundary of a q-simplex, a $(q-1)$-sphere. A polyhedron for which the star of every vertex of a given subdivision is a q-element is called a q-manifold: but in what follows "q-manifold" will mean "connected q-manifold." If M^q is an orientable q-manifold, we denote by $\mathbf{M}^q$ the oriented manifold obtained by orienting

M^q in one of the two possible ways; if $\mathbf{M}^q$, $\mathbf{N}^q$ are both oriented q-manifolds and if ϕ: $M^q \to N^q$ is a *PLO* we shall write ϕ: $\mathbf{M}^q \to \mathbf{N}^q$ if ϕ and the orientation of $\mathbf{M}^q$ induce in N^q the orientation of $\mathbf{N}^q$. If ϕ: $M^q \to M^q$ is an orientation preserving *PLO* we call it a $+PLO$.

In the following sections all theorems will be stated without proof; a full account will be published elsewhere.

2. By R^n we denote Euclidean n-space. The set of $+PLOs$ ϕ: $R^n \to R^n$ form a group of transformations which we denote by $\mathfrak{G}^n$; by $\mathfrak{G}_0^n$ we denote the (normal) subgroup of $\mathfrak{G}^n$ consisting of those $\phi \in \mathfrak{G}^n$ which leave points outside some finite region of R^n invariant (i.e., there must be some such region for each $\phi \in \mathfrak{G}_0^n$).

If $\mathbf{P}$, $\mathbf{Q}$ are (oriented or unoriented) polyhedra of R^n we say that $\mathbf{P}$, $\mathbf{Q}$ are congruent in R^n, $\mathbf{P} \equiv \mathbf{Q}$ in R^n, if there is $\phi \in \mathfrak{G}^n$ such that $\phi\mathbf{P} = \mathbf{Q}$.

THEOREM A. *Let* $\mathbf{P}$, $\mathbf{Q}$ *be finite* q*-dimensional polyhedra of* R^n, *let* ψ: $\mathbf{P} \to \mathbf{Q}$ *be a given PLO and let* $2q + 2 \leq n$. *Then there is* $\phi \in \mathfrak{G}^n$ *such that* $\phi | P = \psi$.

Thus, in particular, equivalent q-dimensional finite polyhedra of R^n are congruent in R^n if $2q + 2 \leq n$.

THEOREM B.[1] *Let* $\mathbf{P}$, $\mathbf{Q}$ *be finite polyhedra of* R^n, *and let* $\phi \in \mathfrak{G}^n$ *be such that* $\phi\mathbf{P} = \mathbf{Q}$. *Then there is* $\phi_0 \in \mathfrak{G}_0^n$ *such that* $\phi_0 | P = \phi | P$.

3. We say that a q-element (a q-sphere) of R^n is flat in R^n if it is congruent in R^n to a q-simplex (the boundary of a $(q + 1)$-simplex).

THEOREM C. *Let* E^q *be a* q*-element of* R^n *and* K *a given simplicial subdivision of* E^q. *Then* E^q *is flat in* R^n *if and only if the star of every vertex of* K *is flat in* R^n.

For $q < n$, we denote by $[q, n]$ the set of congruence classes of oriented q-elements of R^n; for $q < n - 1$ we denote by (q, n) the set of congruence classes of oriented q-spheres of R^n. All n-elements of R^n are flat,[2] while $(n - 1)$-spheres in R^n are the subject of a well-known unproved conjecture.

THEOREM D.[3] $[q, n]$ *and* (q, n) *are commutative (additive) associative systems with a zero element which is the congruence class of flat elements or spheres, respectively.*

The result is also true for $(n - 1)$-spheres in n-space, provided a suitable convention is adopted regarding orientation.

The operation of "addition" implied in Theorem D is obtained as follows in the case $[q, n]$:

Let $\mathbf{E}^q$, $\mathbf{F}^q$ be disjoint oriented q-elements of R^n. Let $\mathbf{g}^q$ be a flat oriented q-element of R^n such that

$$E^q \cap g^q = \dot{E}^q \cap \dot{g}^q = e^{q-1}, \qquad \text{a } q-1\text{-element}$$
$$F^q \cap g^q = \dot{F}^q \cap g^q = f^{q-1}, \qquad \text{a } q-1\text{-element}$$

($\dot{E}^q$ denotes the boundary of E^q, etc.). Furthermore, let e^{q-1}, f^{q-1} have flat q-dimensional neighborhoods in $E^q \cup F^q \cup g^q$, and let the orienta-

tions of $\mathbf{g}^q$, $\mathbf{E}^q$ and $\mathbf{g}^q$, $\mathbf{F}^q$ induce opposite orientations in e^{q-1}, f^{q-1}, respectively. Then $\mathbf{G}^q = \mathbf{E}^q \cup \mathbf{F}^q \cup \mathbf{g}^q$ is an oriented q-element; and we prove that the congruence-class in R^n, γ say, of $\mathbf{G}^q$ depends only on the congruence classes α, β of $\mathbf{E}^q$, $\mathbf{F}^q$. We then define $\alpha + \beta = \beta + \alpha = \gamma$.

A similar construction is used in the case (q, n), a slight further complication being due to the necessity of avoiding a possible "linking" of spheres.

The operation of removing a q-element from the interior of a flat q-element of a q-sphere of R^n leads to a homomorphism

$$\tau\colon (q,n) \to [q, n].$$

The following semigroups are known to be zero:

(q, n) if $2q + 2 \leq n$, by Theorem A.

$[q, n]$ if $2q + 1 \leq n$, by Theorems A and C, using the theory of section 5.

$(1, 2)$, $(2, 3)$,[4] $[1, n]$, $[3, 4]$.

4. Results on embedding in R^n can be applied to embedding in general n-manifolds. The definition of congruence in an n-manifold M^n is the same as that given above in the special case R^n.

Let $\mathbf{P}$ be a polyhedron embedded in an n-manifold M^n. We say that $\mathbf{P}$ is locally embedded in M^n if

(i) There is an n-element $E^n \subset M^n$ such that $\mathbf{P}$ is in the interior of E^n;

(ii) $\mathbf{P}$ does not disconnect M^n.

In particular, these conditions are satisfied if $\mathbf{P}$ is an element in the interior of M^n, or if $\mathbf{P}$ is a sphere in the interior of M^n and $\dim \mathbf{P} < n - 1$.

By $[\mathbf{P}, \mathbf{M}^n]$ we denote the pair consisting of a polyhedron $\mathbf{P}$ locally embedded in an orientable manifold M^n and the oriented manifold $\mathbf{M}^n$.

Let $[\mathbf{P}, \mathbf{M}^n]$, $[\mathbf{Q}, \mathbf{N}^n]$ be two such pairs, let $E^n \subset M^n$ and $F^n \subset N^n$ be n-elements such that $\mathbf{P}$, $\mathbf{Q}$ lie in the interior of E^n, F^n, respectively, and let $\mathbf{E}^n$, $\mathbf{F}^n$ have the orientations induced by $\mathbf{M}^n$, $\mathbf{N}^n$, respectively. If there is a $+PLO$

$$\theta\colon \mathbf{E}^n \to \mathbf{F}^n$$

such that

$$\theta\mathbf{P} = \mathbf{Q},$$

then we say that the pairs $[\mathbf{P}, \mathbf{M}^n]$ and $[\mathbf{Q}, \mathbf{N}^n]$ are congruent, $[\mathbf{P}, \mathbf{M}^n] \equiv [\mathbf{Q}, \mathbf{N}^n]$.

THEOREM E. *Let $\mathbf{P}_i$ $(i = 1, 2)$ be locally embedded in the orientable n-manifold M^n, and $\mathbf{Q}_i$ $(i = 1, 2)$ in the orientable n-manifold $\mathbf{N}^n$. Let $\mathbf{P}_1 \equiv \mathbf{P}_2$ in M^n and $[\mathbf{P}_i, \mathbf{M}^n] \equiv [\mathbf{Q}_i, \mathbf{N}^n]$ $(i = 1, 2)$. Then $\mathbf{Q}_1 \equiv \mathbf{Q}_2$ in N^n.*

COROLLARY 1. *Let M^n, N^n be equivalent manifolds. Then $[\mathbf{P}, \mathbf{M}^n] \equiv [\mathbf{Q}\mathbf{N}^n]$ if and only if there is a PLO $\phi\colon \mathbf{M}^n \to \mathbf{N}^n$ such that $\phi\mathbf{P} = \mathbf{Q}$.*

COROLLARY 2. *$[\mathbf{P}, \mathbf{M}^n] \equiv [\mathbf{Q}, \mathbf{M}^n]$ if and only if $\mathbf{P} \equiv \mathbf{Q}$ in M^n.*

Using this theorem it is easy to set up a one-one correspondence between

congruence classes of oriented polyhedra in an oriented n-manifold and in R^n. Using such a correspondence, for which a definite orientation of R^n must be chosen once and for all, a locally embedded oriented q-sphere in an oriented n-manifold $\mathbf{M}^n$ defines a unique element of (q, n). This remark will be used in the next section.

5. Let $\mathbf{R}^n$ be R^n with a given orientation, and $\mathbf{M}^q$ an oriented q-manifold of R^n, $q < n - 1$. Let $x \in M^q$; we can always find a simplicial subdivision K of R^n of which L, a subdivision of M^q, is a subcomplex and which has x as a vertex. By $\mathbf{St}_L x$ we denote the star of x in L with the orientation induced by $\mathbf{M}^q$ and by $\mathbf{Lk}_L x$ the link of x, i.e., the set of simplices of $\mathbf{St}_L x$ which do not contain x, with the orientation induced by that of $\mathbf{St}_L x$. Similar notations are appropriate to K. Then, using the remark at the end of 4, $[\mathbf{St}_L x, \mathbf{R}^n]$ defines an element $\sigma(x, \mathbf{M}^q)$ of $[q, n]$ and $[\mathbf{Lk}_L x, \mathbf{Lk}_K x]$ defines an element $\lambda(x, \mathbf{M}^q)$ of $[q - 1, n - 1]$ if $x \in M^q$ and of $(q - 1, n - 1)$ if x is interior to M^q. The elements $\sigma(x, \mathbf{M}^q)$ and $\lambda(x, \mathbf{M}^q)$ are independent of the partitions used in their definition. $\sigma(x, \mathbf{M}^q) = 0$ if and only if $\lambda(x, \mathbf{M}^q) = 0$.

If $n = 2q$, it is a consequence of Theorem A that $\lambda(x, \mathbf{M}^q) = 0$ unless x is a vertex of some given partition of M^q; hence, for a finite manifold, there is at most a finite set of points $x \in M^q$ such that $\lambda(x, \mathbf{M}^q) \neq 0$; also these points must be interior ones. Thus, with such a manifold we can associate uniquely a certain finite set of elements of $(q - 1, n - 1)$, called the "link type" of $\mathbf{M}^q$.

THEOREM F. *The link type establishes a one-one correspondence between the elements of $[q, 2q]$ and (unordered) sets of the form $(\beta_1, \ldots, \beta_k)$ where $\beta_i \epsilon (q - 1, 2q - 1)$, where the empty set stands for the zero element, and addition is given by $(\beta_1, \ldots, \beta_k) + (\beta_{k+1}, \ldots, \beta_l) = (\beta_1, \ldots, \beta_l)$.*

6. The two theorems of this section are used in the proofs of the theorems on embedding:

Let P, Q be polyhedra and each of ϕ_0, ϕ_1: $P \to Q$ a *PLO*. If there is a *PLO* ϕ_t: $P \times I \to Q \times I$, where I denotes the unit interval and $P \times I$ the topological product, such that $\phi_t(p, t) = (q, t)$, $\phi_t(p, 0) = (\phi_0 p, 0)$ and $\phi_t(p, 1) = (\phi_1 p, 1)$ for $p \in P$, $t \in I$, then we say that ϕ_0 and ϕ_1 are (piecewise onto) isotopic, and we write $\phi_0 \backsimeq \phi_1$.

THEOREM G.[5] *Let P be either an element or a sphere, and ϕ: $P \to P$ a $+PLO$. Then $\phi \backsimeq 1$ (the identity map).*

Let M^q be an orientable q-manifold, and E_i^q, $i = 1, 2$, q-elements in its interior. Let $P \subset M^q - E_1^q - E_2^q$ be a polyhedron which does not disconnect M^q, and let ψ: $E_1^q \to E_2^q$ be a *PLO* which is orientation preserving in M^q.

THEOREM H.[6] *There is a $+PLO$ ϕ: $M^q \to M^q$ such that*

$$\phi | P = 1, \qquad \phi | E_1^q = \psi, \qquad \phi \backsimeq 1.$$

[1] This result was previously obtained, by a method differing from ours, by Graeub, W., "Die Semilinearen Abbildungen," *Sitzungsber. Heidelberger Akad. Wissensch.*, 1950, § 3, Satz 2.

[2] Newman, M. H. A., "On the Superposition of n-Dimensional Manifolds," *J. London Math Soc.*, 1927.

[3] The associative system (1, 3) was discovered by Schubert, H., "Die Eindeutige Zerlegunge ines Knotens in Primknoten," *Sitzungsber. Heidelberger Akad. Wissensch.*, 1949.

[4] Graeub, W., *loc. cit.*, § 5, Satz 1.

[5] A similar theorem was proved for the case when P is an element and $\phi|\dot{P} = 1$ by Veblen, O., "On the Deformation of an n-Cell," these PROCEEDINGS, **3**, 1917, and for the case P a sphere and ϕ a differentiable map by Whitehead, J. H. C. (unpublished; I am indebted to Prof. Whitehead who told me of his results). Both these results concerned the classical definition of isotopy.

[6] This extends a result due to Newman, M. H. A., *loc. cit.*, and strengthened by Whitehead, J. H. C., "On Subdivisions of a Complex," *Cambridge Phil. Soc. Proc.*, 1935.

PLANCHEREL FORMULA FOR THE 2 × 2 REAL UNIMODULAR GROUP

BY HARISH-CHANDRA

DEPARTMENT OF MATHEMATICS, COLUMBIA UNIVERSITY.

Communicated by P. A. Smith, February 27, 1952

In an earlier note[1] in these PROCEEDINGS we have obtained an *explicit* Plancherel formula for a connected complex semisimple Lie group. The corresponding problem for a real group appears to be much more difficult due to the circumstance that two Cartan subgroups in it are in general not conjugate. In fact it seems likely that there is a close connection between classes of conjugate Cartan subgroups and the various "series" of unitary representations which occur in the Plancherel formula. The object of this note is to make a beginning in the study of this question by deriving an *explicit* Plancherel formula in the simple case of 2×2 real unimodular group.[2]

Let R be the field of real numbers and G the group of all 2×2 real matrices with determinant 1. Then its Lie algebra $\mathfrak{g}_0$ consists of all matrices with trace zero. Put

$$H = \begin{pmatrix} 1/2 & 0 \\ 0 & -1/2 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad U = \begin{pmatrix} 0 & 1/2 \\ -1/2 & 0 \end{pmatrix}$$

and

$$h_t = \exp tH, \quad n_s = \exp sX, \quad u_\theta = \exp \theta U \qquad (t, s, \theta \in R).$$

Let K and A be the one parameter subgroups G corresponding to U and

H, respectively. Then the mapping $(u, t, s) \to uh_t n_s (u \in K, t, s \in R)$ is a topological mapping of $K \times R \times R$ onto G. Let G^* denote the adjoint group of G and $x \to x^*$ $(x \in G)$ the natural mapping of G on G^*. We denote by η_j the character of K given by

$$\eta_j(u_\theta) = e^{ij\theta} \qquad (j = 0, 1/2).$$

(i is a fixed square-root of -1). Let du^* denote the element of Haar measure on K^* such that $\int_{K^*} du^* = 1$ and let $\mathfrak{H}$ be the Hilbert space of all square-integrable functions on K^*. Then for any complex number ν we define a representation $\pi_{\nu, j}$ $(j = 0, 1/2)$ of G on $\mathfrak{H}$ as follows. If $x \in G$ and $u \in K$ we write

$$xu = u_x \exp(t(x, u)H) \exp(s(x, u)X)$$

where $u_x \in K$ and $t(x, u)$, $s(x, u) \in R$. It is easy to verify that for fixed x, $(u_x)^*$, $u^{-1}u_x$ and $t(x, u)$ depend only on u^* and so we may denote them by u_x^*, $v(x, u^*)$ and $t(x, u^*)$, respectively. Now for any $f \in \mathfrak{H}$ we define $\pi_{\nu, j}(x)f = g$ by the rule

$$g(u^*) = \eta_j(v(x^{-1}, u^*))e^{-\nu t(x^{-1}, u^*)} f(u^*_{x^{-1}}).$$

It can be checked without difficulty that $\pi_{\nu, j}$ is actually a representation and if $\nu + 1/2$ is purely imaginary $\pi_{\nu, j}$ is unitary. Put $\tau_\lambda^+ = \pi_{\nu, 0}$ and $\tau_\lambda^- = \pi_{\nu, 1/2}$ where $\nu = i\lambda - 1/2 (\lambda \in R)$. Then τ_λ^+ and τ_λ^- are unitary representations of G on $\mathfrak{H}$.

Now consider the representation $\pi_{n+j, j}$ where n is an integer ≥ 0. For any integer m let f_m denote the function $f_m(u_\theta^*) = e^{im\theta}$. Then these functions form an orthonormal base for $\mathfrak{H}$. Let $V_{n, j}$ be the subspace of $\mathfrak{H}$ spanned by the set $\{f_m;\ |m + j| \leq n + j\}$ and let $W_{n, j}$ be the orthogonal complement of $V_{n, j}$ in $\mathfrak{H}$. Then $V_{n, j}$, $W_{n, j}$ are invariant subspaces and we denote by $\rho_{n, j}$, $\sigma'_{n, j}$ the corresponding representations of G induced on them. It is known that there exists a unitary representation $\sigma_{j, n}$ which is infinitesimally[3] equivalent to $\sigma'_{j, n}$. Let $\chi_{n, j}$ be the character of $\rho_{n, j}$. Then

$$\left.\begin{aligned} \chi_{n, j}(h_t) &= \frac{e^{(n+j+1/2)t} - e^{-(n+j+1/2)t}}{e^{1/2t} - e^{-1/2t}} \\ \chi_{n, j}(u_\theta) &= \frac{\sin(n + j + 1/2)\theta}{\sin\dfrac{\theta}{2}} \end{aligned}\right\} \qquad (1)$$

Let $\gamma = \exp(2\pi U)$. For any function f on G put $f_+(x) = f(x) + f(\gamma x)$ and $f_-(x) = f(x) - f(\gamma x)$. Let $C_c^\infty(G)$ denote the set of all functions on G which vanish outside a compact set and which are indefinitely differentiable. Put

$$T_\lambda^*(f) = \int f_*((h_t n_s)^{u^*}) e^{(i\lambda + 1/2)t}\, dt\, ds\, du^* \qquad (2)$$

and

$$S_k^*(f) = \int f_*((h_t n_s)^{u^*}) e^{(k+1)t}\, dt\, ds\, du^* - \int_G f_*(x)\chi_k(x)\, dx \quad (3)$$

Here $y^{x^*} = xyx^{-1}$ ($x, y \in G$) and $dx = e^t\, dt\, ds\, du$ ($x = uh_t n_s$), du being so normalized that $\int_K du = 1$. $2k$ is an integer ≥ 0 and $\chi_k = \chi_{k,0}$ or $\chi_{k-1/2,\,1/2}$ according as $2k$ is even or odd and we have to take the plus or minus sign in (3) accordingly. It can be seen without difficulty[4] that $T_\lambda^*(f)$, $S_k^*(f)$, respectively, are the traces of the operators $\int_G f(x)\tau_\lambda^*(x)\, dx$ and $\int_G f(x)\sigma_k^*(x)\, dx$ where $\sigma_k^+ = \sigma_{k,0}$ or $\sigma_k^- = \sigma_{k-1/2,\,1/2}$ according as $2k$ is even or odd. Moreover since the conjugates of K and A taken together cover G except for a set of measure zero it follows easily from (1) that

$$S_k^*(f) = \int f_*((h_t n_s)^{u^*}) e^{t/2} e^{-(k+1/2)|t|}\, dt\, ds\, du^* - \frac{2}{\pi}\int_{-\pi}^{\pi} d\theta \sin\left(k+\frac{1}{2}\right)\theta \sin\frac{\theta}{2}\int_{G^*} f_*((u_\theta)^{x^*})\, dx^* \quad (4)$$

where $dx^* = e^t\, dt\, ds\, du^*$ ($x^* = u^*(h_t n_s)^*$).

THEOREM. *Let $f \in C_c^\infty(G)$. Then*[5]

$$8\pi f(1) = \frac{1}{2}\int_{-\infty}^{\infty} \lambda \tanh \pi\lambda\, T_\lambda^+(f)\, d\lambda + \frac{1}{2}\int_{-\infty}^{\infty} \lambda \coth \pi\lambda\, T_\lambda^-(f)\, d\lambda + \sum_{n\geq 0}\left(n+\frac{1}{2}\right) S_n^+(f) + \sum_{n\geq 0}(n+1)S_{n+1/2}^-(f)$$

the two series being absolutely convergent.

The proof is based on the following lemma.

LEMMA. *Let $F(x_1, x_2)$ be an indefinitely differentiable function of two real variables which vanishes outside a compact set. Put*

$$g(\theta) = \int_0^\infty (e^t - e^{-t})F(\theta e^t, \theta e^{-t})\, dt \qquad (\theta > 0)$$

Then

$$\lim_{\theta\to 0} \theta g(\theta) = \int_0^\infty F(s, 0)\, ds$$

$$\lim_{\theta\to 0} \frac{d}{d\theta}(\theta g(\theta)) = -2F(0, 0)$$

and there exist positive constants a and b such that

$$\left|\frac{1}{\theta}\left(\frac{d}{d\theta}(\theta g(\theta)) + 2F(0, 0)\right)\right| \leq a + b|\log\theta|$$

for $0 < \theta \leq 1$.

In proving the theorem we may assume that $f(uxu^{-1}) = f(x) = f(x^{-1})$ $(x \in G, u \in K)$. Put

$$F_{\pm}(\theta) = \int_{G^*} f_{\pm}((u_\theta)^{x^*})\, dx^*.$$

Then making use of the above lemma we find that

$$\operatorname*{Lim}_{m \to \infty} \left\{(m+1)\int_{-\infty}^{\infty} f_+(n_s)\, ds - \sum_{0 \le n \le m}\left(n+\frac{1}{2}\right)\frac{1}{\pi} \times \int_{-\pi}^{\pi} \sin\left(n+\frac{1}{2}\right)\theta \sin\frac{\theta}{2} F_+(\theta)\, d\theta\right\} = -\operatorname*{Lim}_{m \to \infty} \sum_{0 \le n \le m} \frac{1}{\pi} \times \int_{-\pi}^{\pi} \cos\left(n+\frac{1}{2}\right)\theta \frac{d}{d\theta}\left(\sin\frac{\theta}{2} F_+(\theta)\right) d\theta = 2\pi f_+(1).$$

Now put $\varphi_{\pm}(t) = e^{t/2}\int_{-\infty}^{\infty} f_{\pm}((h_t n_s)\, ds$. Then

$$\sum_{0 \le n \le m}\left(n+\frac{1}{2}\right)\int_0^{\infty} \varphi_+(t) e^{-(n+1/2)t}\, dt = (m+1)\varphi_+(0) + \sum_{0 \le n \le m}\int_0^{\infty} \frac{d\varphi_+(t)}{dt}\, e^{-(n+1/2)t}\, dt.$$

Hence

$$4\pi f_+(1) = \operatorname*{Lim}_{m \to \infty}\left\{\sum_{0 \le n \le m}\left(n+\frac{1}{2}\right) S_n^+(f) - \sum_{0 \le n \le m} 2\int_0^{\infty} \frac{d\varphi_+(t)}{dt}\, e^{-(n+1/2)t}\, dt\right\}.$$

Since $\varphi_+(t) = \varphi_+(-t)$, $\dfrac{1}{t}\dfrac{d\varphi_+(t)}{dt}$ is a differentiable function of t and therefore

$$\operatorname*{Lim}_{m \to \infty} \sum_{0 \le n \le m}\int_0^{\infty} \frac{d\varphi_+(t)}{dt}\, e^{-(n+1/2)t}\, dt = \int_0^{\infty} \frac{1}{t}\frac{d\varphi_+(t)}{dt}\frac{t}{(e^{t/2}-e^{-t/2})}\, dt = -\frac{1}{4}\int_{-\infty}^{\infty} \lambda \tanh \pi\lambda\, T_\lambda^+(f)\, d\lambda$$

by Fourier transformation since $T_\lambda^+(f) = \int_{-\infty}^{\infty} \varphi_+(t) e^{i\lambda t}\, dt$. Hence

$$4\pi f_+(1) = \frac{1}{2}\int_{-\infty}^{\infty} \lambda \tanh \pi\lambda\, T_\lambda^+(f)\, d\lambda + \sum_{0 \le n < \infty}\left(n+\frac{1}{2}\right) S_n^+(f). \quad (5)$$

Similarly we prove that

$$\sum_{n \geq 0} (n + 1)\, S^{-}_{n+1/2}(f) = 4\pi f_{-}(1) + \int_0^\infty \frac{d\varphi_{-}(t)}{dt} \coth \frac{t}{2}\, dt =$$

$$4\pi f_{-}(1) - \frac{1}{2} \int_{-\infty}^{\infty} \lambda \coth \pi\lambda\, T_\lambda^{-}(f)\, d\lambda. \quad (6)$$

Adding (5) and (6) we get the result. The Plancherel formula for G can now be obtained in the usual way.

Notice that if $f(uxv) = f(x)$ $(u, v \in K)$ then $T_\lambda^{-}(f) = S_n^{+}(f) = S^{-}_{n+1/2}(f) = 0$ and therefore

$$8\pi f(1) = \frac{1}{2} \int_{-\infty}^{\infty} \lambda \tanh \pi\lambda\, T_\lambda^{+}(f)\, d\lambda. \quad (7)$$

If we assume, as we may without essential loss of generality, that $f(x) = f(x^{-1})$ $(x \in G)$ and put

$$\varphi(t) = e^{t/2} \int_{-\infty}^{\infty} f(h_t n_s)\, ds$$

then (7) is equivalent to

$$2\pi f(1) = -\int_{-\infty}^{\infty} \frac{d\varphi(t)}{dt} (e^{t/2} - e^{-(t/2)})^{-1}\, dt.$$

This formula is quite easy to verify directly if we take into account the fact that the function $F(t, s) = f(h_t n_s)$ depends only on $(1 + s^2)e^t + e^{-t} = (e^{t/2} - e^{-(t/2)})^2 + s^2 e^t - 2$ which is the trace of the product of the matrix $h_t n_s$ with its transposed. Put $x = (e^{t/2} - e^{-(t/2)})$, $y = se^{t/2}$ and $\Phi(r) = F(t, s)$ where $r = \sqrt{x^2 + y^2} \geq 0$. Then

$$\varphi(t) = \int_{-\infty}^{\infty} \Phi(r)\, dy$$

and therefore

$$\frac{d\varphi(t)}{dt} = \frac{dx}{dt} \int_{-\infty}^{\infty} \frac{x}{r} \frac{d\Phi(r)}{dr}\, dy.$$

Hence

$$\int_{-\infty}^{\infty} \frac{d\varphi(t)}{dt} (e^{t/2} - e^{-(t/2)})^{-1}\, dt = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{r} \frac{d\Phi(r)}{dr}\, dx\, dy =$$

$$2\pi \int_0^\infty \frac{d\Phi(r)}{dr}\, dr = -2\pi\Phi(0) = -2\pi f(1).$$

Formula (7) has also been obtained independently by R. Godement.

[1] PROC. NATL. ACAD. SCI., **37**, 813–818 (1951).

[2] The representations of this group have been studied by Bargmann, *Ann. Math.*, **48**, 568–640 (1947).

[3] For the definition of infinitesimal equivalence see an earlier note (PROC. NATL. ACAD. SCI., **37**, 362–365 (1951)). $\sigma_{j,n}$ is the direct sum of two irreducible representations which in Bargmann's notation are D^{+}_{n+1+j} and D^{-}_{n+1+j}.

[4] See PROC. NATL. ACAD. SCI., **37**, 366–369 (1951).

[5] Compare this result with formulas (11.7), (12.12) and (12.24) of Bargmann's paper.

*ON SOLUTIONS OF MAXWELL'S EQUATIONS IN AN EXTERIOR REGION**

BY WILLIAM K. SAUNDERS

UNIVERSITY OF CALIFORNIA, BERKELEY

Communicated by H. Weyl, February 2, 1952

1. This paper is concerned with the questions of the existence and uniqueness of solutions to the time periodic case of Maxwell's equations for an exterior region. The work is limited to constant real parameters. The results are as follows:

THEOREM I. *Given an infinite region L surrounding a finite closed smooth surface C with an assignment upon C of the tangential components of a vector field* $\mathbf{E}$ *of class* C^2 *with respect to the surface parameters of C, there exists at most one set of twice continuously differentiable time periodic functions satisfying Maxwell's equations in L, the boundary conditions upon* $\mathbf{E}$ *on C, and the condition*[1]

$$\lim_{r \to \infty} r(\mathbf{a}_r \times \mathbf{H} + \sqrt{\epsilon/\mu}\,\mathbf{E}) = 0 \qquad (1)$$

THEOREM II.—*There exists in L a solution of the time periodic Maxwell's equations satisfying the boundary and infinity conditions of Theorem I.*

The proof of the first theorem is an extension of the proof of a theorem of F. Rellich[2] upon solutions of the scalar Helmholtz equation. The second theorem is deduced from the observation that the proof is contained inherently in a recent result of C. Müller[3] upon the problem of a perfect conductor.

In 1912–1914 H. Weyl[4] in a series of papers on the asymptotic distribution of the eigenvalues of various partial differential equations dealt with the electromagnetic problem in a region enclosed by a perfect conductor. To this end he set up the vector Green's formula,

$$\int_L \{\mathbf{E}\cdot[\nabla^2 G] - [\nabla^2\mathbf{E}]\cdot G\}\,dp = \int_C \{-[\mathbf{n}\times\mathbf{E}]\cdot[\nabla\times G] + [\nabla\times\mathbf{E}]\cdot[\mathbf{n}\times G] - [\mathbf{E}\cdot\mathbf{n}]\cdot[\nabla\cdot G] + [\nabla\cdot\mathbf{E}]\cdot[\mathbf{n}\cdot G]\}\,do \qquad (2)$$

where G may be a dyadic or tensor as well as a vector. A method, involv-

ing the solution of a system of integral equations, was given to construct a tensor G which would satisfy the vector Laplace equation. F. H. Murray[5] and the author[6] attempted to vary the procedure to construct a tensor which would satisfy the vector Helmholtz equation hoping thereby, through the use of (2), to write the field in terms of its assigned boundary values. Although the method of Weyl requires little modification and the eigenvalues situation may be handled formally in a parallel manner, the work is none too satisfactory. In the first place, there remains a question whether the solution constructed is truly one for Maxwell's equations and not merely one for the larger class belonging to the vector Helmholtz equation, and in the second place, the method used for consideration of the eigensolutions lacks rigor especially when applied to other than the Laplace case.

In a recent paper H. Weyl[7] has given a far more satisfactory method for handling the eigensolutions. He limits himself there to the scalar Helmholtz equation but has indicated that the method may be applied with equal success to Maxwell's equations.[8] This is especially true if the construction of the tensor G is done in the manner of Weyl's Palermo[9] paper which has the advantage of overcoming the first difficulty mentioned above. With this method there is no question but that the field constructed is divergenceless. At the time of the publication of C. Müller's paper the author was revising his paper to prove Theorem II in the manner just indicated. It seems best, however, to adopt the simpler derivation of Müller and to show that this solution to the reflection problem is in essence all that is required.[10]

2. THEOREM (Rellich). Let ϕ for $\rho > \rho_0$ be a twice continuously differentiable complex valued solution of $(\nabla^2 + k^2)\phi = 0$ and let ϕ be not identically equal to 0, then there exists a $t > 0$ such that for all sufficiently large r (and an arbitrary $\rho_1 > \rho_0$)

$$\int\int\int_{\rho_1 < \rho < r} |\phi|^2 \, dp \geq tr$$

(dp = element of volume). In particular one has the consequence:

$$\text{If } \lim_{r \to \infty} \int\int_{\Sigma} |\phi|^2 \, d\Omega = 0 \qquad \text{then } \phi = 0$$

(Σ = a spherical surface of radius r; Ω = the solid angle.)

COROLLARY (Rellich). *There is at most one complex valued function which in the region L is twice continuously differentiable, assumes given values on C, satisfies in L the equation*

$$(\nabla^2 + k^2)\phi = 0$$

and the condition

$$\lim_{r \to \infty} r \left(\frac{\partial \phi}{\partial r} + ik\phi \right) = 0 \tag{3}$$

uniformly for all directions $\mathbf{r}$. *The values of* ϕ *assigned upon* C *must be such that Green's theorem is applicable.*

If this corollary is applied in a trivial manner it is apparent that if each of the three rectangular components of $\mathbf{E}$ satisfies the Sommerfeld radiation condition (3) and if all three components are independently given upon a boundary C, then the electromagnetic field is uniquely determined in the exterior region. There are two objections to a uniqueness proof in this form. The first is that usually only the tangential components of the electric field are known on C and it is not immediately apparent that these tangential components determine the normal components although, with the completion of a uniqueness theorem based only on knowledge of the tangential components, such is implied. A second objection is that to assume the Sommerfeld condition for each of the three rectangular components of $\mathbf{E}$ is to assume it in particular for the radial component E_r and although one might expect the special behavior called for in the Sommerfeld condition with the the E_θ and E_ϕ components it is common knowledge that this is not a natural condition to set upon the E_r components. From the physical point of view one would expect that at each point on a large sphere surrounding C the radiation has the property of an outgoing plane wave. This is expressed mathematically by (1) above.[11]

Proof: Theorem I—Suppose there exist two solutions E' and E'', with corresponding H' and H'', satisfying Maxwell's equations and (1), and form

$$\mathbf{W} = \mathbf{E}' - \mathbf{E}'' \qquad \mathbf{Y} = \mathbf{H}' - \mathbf{H}''$$
$$\mathbf{W}^* = \mathbf{W} \text{ Conjugate} \qquad \mathbf{Y}^* = \mathbf{Y} \text{ Conjugate}$$

Set $\mathbf{E} = \mathbf{W}$, $G = \mathbf{W}^*$ in (2) (the divergence terms vanish), note that $\mathbf{W}$, $\mathbf{Y}$, $\mathbf{W}^*$, $\mathbf{Y}^*$ satisfy equations like Maxwell's equations and make use of this in the right-hand term. Then

$$\int_L \{\mathbf{W}^* \cdot \nabla \times \nabla \times \mathbf{W} - \mathbf{W} \cdot \nabla \times \nabla \times \mathbf{W}^*\}\, dp = \int_{C+\Sigma} \{\mathbf{W} \times i\omega\mu\mathbf{Y}^* + \mathbf{W}^* \times i\omega\mu\mathbf{Y}\} \cdot n\, do$$

Here C is a finite surface, Σ a large spherical surface near ∞. The volume L lies between them. The usual manipulation yields the vector Helmholtz equation without divergence for $\mathbf{E}$ and hence for $\mathbf{W}$ and $\mathbf{W}^*$. In the source-free region L, therefore, the integrand of the left side reduces to $\mathbf{W}^* \cdot k^2\mathbf{W} - k^2\mathbf{W} \cdot \mathbf{W}^* = 0$ and the integral vanishes. Since by the hypotheses on C, $E'_{\tan} = E''_{\tan}$, $W_{\tan} = W^*_{\tan} = 0$ and since W_{normal} gives no contribution, the integral over C also vanishes and thus

$$i\omega\mu \int_\Sigma (\mathbf{W} \times \mathbf{Y}^* + \mathbf{W}^* \times \mathbf{Y}) \cdot \mathbf{a}_r\, do = 0$$

Expressing condition (1) in terms of $\mathbf{Y}$ and $\mathbf{W}$, and also in terms of $\mathbf{Y}*$ and $\mathbf{W}*$, one has

$$\lim_{r \to \infty} r^2(\mathbf{a}_r \times \mathbf{Y} + \sqrt{\epsilon/\mu}\mathbf{W}) \cdot (\mathbf{a}_r \times \mathbf{Y}^* + \sqrt{\epsilon/\mu}\mathbf{W}^*) = 0$$

and if one integrates over Σ

$$\lim_{r \to \infty} \int_\Sigma \{(\mathbf{Y}_{\tan} \cdot \mathbf{Y}^*_{\tan}) + \sqrt{\epsilon/\mu}(\mathbf{Y} \times \mathbf{W}^* + \mathbf{Y}^* \times \mathbf{W}) + \epsilon/\mu \mathbf{W} \cdot \mathbf{W}^*\} d\Omega = 0$$

Since the second term vanishes and the integrand of the remaining terms is the sum of positive squares

$$\lim_{r \to \infty} \int_\Sigma |W_x|^2 \, d\Omega = 0 \qquad \text{etc.}$$

But since W_x, W_y, and W_z are rectangular components of the difference field satisfying $(\nabla^2 + k^2)W_x = 0$ by the Theorem of Rellich above, $W_x = 0$, $W_y = 0$, $W_z = 0$. Hence $\mathbf{W} = 0$ and $\mathbf{E}' = \mathbf{E}''$ which was to be shown.

3. Although it should be possible to show directly that a field such as the one to be constructed in the proof of Theorem II meets the condition (1) above, an alternative demonstration making use of a triplet of Sommerfeld conditions proves more convenient. To this end one writes down the condition (1) in terms of the electric field in spherical coordinates.

$$(a) \quad \lim_{r \to \infty} r\left[\frac{1}{i\omega\mu r} E_\theta + \frac{1}{i\omega\mu} \frac{\partial E_\theta}{\partial r} - \frac{1}{i\omega\mu r} \frac{\partial E_r}{\partial \theta} + \sqrt{\epsilon/\mu}\, E_\theta\right] = 0$$

$$(b) \quad \lim_{r \to \infty} r\left[\frac{-1}{i\omega\mu r \sin\theta} \frac{\partial E_r}{\partial \phi} + \frac{E_\phi}{i\omega\mu r} + \frac{1}{i\omega\mu} \frac{\partial E_\phi}{\partial r} + \sqrt{\epsilon/\mu}\, E_\phi\right] = 0 \qquad (4)$$

$$(c) \quad \lim_{r \to \infty} rE_r = 0$$

and for comparison the triplet of Sommerfeld conditions taken also in spherical coordinates (i.e., in linear combinations of the set in rectangular coordinates)

$$(a) \quad \lim_{r \to \infty} r\left[\frac{\partial E_\theta}{\partial r} + ikE_\theta\right] = 0$$

$$(b) \quad \lim_{r \to \infty} r\left[\frac{\partial E_\phi}{\partial r} + ikE_\phi\right] = 0 \qquad (5)$$

$$(c) \quad \lim_{r \to \infty} r\left[\frac{\partial E_r}{\partial r} + ikE_r\right] = 0$$

Suppose in addition the field $\mathbf{E}$ is divergenceless, i.e.,

$$\frac{\partial E_r}{\partial r} + \frac{2}{r} E_r + \frac{1}{r} \frac{\partial E_\theta}{\partial \theta} + \frac{1}{r \sin \theta} \frac{\partial E_\phi}{\partial \phi} = 0 \qquad (6)$$

Comparing (4a) with (5a) and (4b) with (5b) one notes that they are equivalent provided that the terms E_θ, $\frac{\partial E_r}{\partial \theta}$, E_ϕ and $\frac{\partial E_r}{\partial \phi}$ vanish as $r \to \infty$ to any order whatever, for the terms appearing in (4a, b) but not in (5a, b) all have an $1/r$ additional factor. (Note that $k = \omega\sqrt{\epsilon\mu}$.) And under the requirement that $\frac{\partial E_\theta}{\partial \theta}$, E_r, $\frac{\partial E_\phi}{\partial \phi}$ all vanish to some order one sees from that $\lim_{r \to \infty} r \frac{\partial E_r}{\partial r} = 0$, and hence learns the equivalence of (4c) and (5c). If then the field $\mathbf{E}$ and its derivatives vanish at large distances from the sources and if $\mathbf{E}$ has no divergence, then the triplet of Sommerfeld conditions in (r, θ, ϕ) or (x, y, z) is equivalent to the condition (1).

4. THEOREM (Müller). There exists a solution to the following problem. Given a perfect conductor coinciding with the surface C, and a set of sources of electromagnetic energy within a finite domain B which does not intersect C, find a solution to Maxwell's equations in the region exterior to B and C which meets the assigned conditions in B, the condition $\mathbf{n} \times \mathbf{E} = 0$ on C and the infinity conditions of Theorem I above.

Müller replaces the conductor by an unknown distribution of magnetic current $\mathbf{j}_0$. The condition $\mathbf{n} \times \mathbf{E} = 0$ on C leads directly to the vector integral equation

$$2\mathbf{j}_e = -2\pi\mathbf{j}_0 - \int_C \mathbf{n} \times (\mathbf{j}_0 \times \nabla\phi)\, do \qquad \phi = e^{-ikR}/R$$

where $\mathbf{j}_e$ is the magnetic current equivalent to the tangential components of the incident electric field upon C. The homogeneous part of this equation and the homogeneous part of the adjoint differ but little. In fact it is shown that if $\mathbf{j}_0'$ is a solution to the first then $\mathbf{t} = -\mathbf{n} \times \mathbf{j}_0'$ is a solution to the second.

With the solutions of the homogeneous equations orthogonal in this vector sense, it is then not difficult to show that $\mathbf{t}$ and $\mathbf{j}_e$ are orthogonal in the manner required for the Fredholm theory. The proof consists of the use of (2) and a limiting process onto C from the interior.

Using this result of Müller it is simple to construct the necessary G for use in (2) and thus obtain a proof of Theorem II. Let G be the sum of three dyadics $A + F + M$. For A take $e^{-ikR}I/R$ where $R = (r - r')$ and I is the idemfactor. Considering each column as a vector field it is apparent that none of these fields is divergenceless, and that to each must be added a second vector field if the sum is to be a solution to Maxwell's equations. What is required is that

$$\nabla \times \nabla \times A(+ F) - k^2(A + F) = \delta(r - r')I \qquad (7)$$

The vector identity $\nabla \times \nabla \times \mathbf{C} = -\nabla^2 C + \nabla\nabla \cdot \mathbf{C}$ and the knowledge that

$$(\nabla^2 + k^2)e^{-ikR}/R = \delta(r - r')$$

adequately suggest the character of F. Thus if $F = 1/k^2\nabla\nabla e^{-ikR}/R$ then it will just cancel out the unwanted term arising from $\nabla\nabla e^{-ikR}/R$ without making a contribution due to the application of the $\nabla \times \nabla \times$ operation. Consider now the matrix $A + F$. Each column is in the form of singular electric field, i.e., a solution of (7) above having suitable behavior at infinity. If $A + F$ is substituted for G in (2) one has, by the use of the Dirac relation (7) above or by the somewhat more tedious exclusion of a small sphere[12] $\mathbf{E} = -\frac{1}{4\pi}\int_C \{-[\mathbf{n} \times \mathbf{E}] \cdot \nabla \times (A + F) + [\nabla \times \mathbf{E}] \cdot [\mathbf{n} \times (A + F)]\}\, do$.

If finally there is added to each column of the matrix $A + F$ the reflected vector field guaranteed by the Müller theorem there arises the sum $A + F + M$ which has the property that on C, $\mathbf{n} \times (A + F + M) = 0$. Thus each column of the matrix $A + F + M$ consists of an incident electromagnetic field (of divergence zero) and its reflection in a perfect conductor coincident with C. Moreover the total field represented by each column has just the singularity noted above. With $A + F + M$ substituted for G one has the desired construction.

$$\mathbf{E} = -\frac{1}{4\pi}\int - (\mathbf{n} \times \mathbf{E}) \cdot (\nabla \times [A + F + M])\, do$$

Acknowledgment.—I wish to thank Professors Evans, Silver and Weyl for suggestions and encouragement.

* Work supported in part by Office of Naval Research.

[1] $\mathbf{a}_r$ is the unit radial vector, r the distance between any source point and the field point.

[2] Rellich, F., *Jahresber d. D. M. V.*, **53,** 57–64 (1943).

[3] Müller, Claus, *Math. Ann.*, **123,** 345–378 (1951); especially pp. 373–378.

[4] Weyl, H., *Ibid.*, **71,** 441–479; *J. Math.*, **143,** 177–202.

[5] Murray, F. H., these PROCEEDINGS, **16,** 353–357 (1930); *Am. J. Math.*, **53,** 275–288 (1931).

[6] Saunders, W. K., Thesis, University of Calif. (1951). Also *Antenna Lab. Rep.* No. 175–176 (1950–1951).

[7] Weyl, H., these PROCEEDINGS, 832 (1951). Detailed version to appear in *Math. Zeit.*

[8] Private communication and brief note at the end of PROC. NATL. ACAD. SCI. paper above.

[9] Weyl, H., *Cir. Mat. di Palermo*, **39,** 1–50 (1915).

[10] The connection between the two problems has been suggested by Schwinger, J. (unpublished but widely distributed notes); Severin, Hans (*Z. Physik*, **129,** 426–439) (1951); and probably many others unknown to the present author.

[11] The mathematical statement of this condition was given independently by Silver, S., *Microwave Antenna Theory and Design*, McGraw-Hill, 1949, section 39 and by Müller, Claus, *Arch. Math.*, **1** (1948–1949). Both noted that it was one of four necessary conditions for the vanishing of the integral at infinity in an existence proof based upon both tangential E and tangential H given by Stratton and Chu, *Phys. Rev.*, **56**, 99 (1939).

[12] Although terms of $F \to (1/R^3)$ as $r \to r'$ it will be found that the integral involving the regular field $\nabla \times \mathbf{E}$ vanishes by consideration of symmetry while $\nabla \times F = 0$ automatically.

---

*ON COHOMOLOGY THEORIES**

By Chung-Tao Yang

Department of Mathematics, Tulane University

Communicated by E. J. McShane, February 29, 1952

It is only[1] recently that the usefulness of fully normal spaces[2] in algebraic topology has been recognized. We note first that this category of spaces contains both metric and compact Hausdorff spaces.[3] Further, A. H. Stone[4] has shown that for Hausdorff spaces full normality is the same as paracompactness. Now it is well known that finiteness conditions (e.g., finite open coverings) lead to non-intuitive results for very simple spaces. In order to avoid this situation (as well as for other reasons) it is customary to introduce compactness in some form, compact supports, compact homologies and so on. This, however, introduces difficulties in applications. Very few function spaces, for example, are provided with a sufficient number of compact subsets. Many of the more interesting ones are metric and hence fully normal. It is desirable to develop a full-fledged homology theory applicable to fully normal spaces and not requiring any compactness conditions. For reasons now familiar the singular theory is inadequate. Even for locally compact connected finite-dimensional groups satisfactory results about regularity in the small have not yet been obtained in sufficient amount to permit application of the singular theory. One is then inclined toward the Čech theory (using quite arbitrary coverings) and toward cohomology rather than homology, since in the former a discrete coefficient group may be used. However, the Alexander-Kolmogoroff theory is more immediate and direct, not requiring the elaborate machinery of complexes, orientation (or ordering) and limit-groups essential to even the definition of the Čech groups. There is the additional advantage that the Eilenberg-Steenrod "axioms" (except the homotopy axiom) are known[5] to be satisfied in this theory with no restrictions at all on the spaces. Much more is known[6] when the space is fully normal. On the other hand Dowker[7] has shown that for the unrestricted Čech groups "the axioms," including the

homotopy axiom, are all valid. He did not, however, prove anything beyond the weak excision theorem.

In this paper it is shown that, for fully normal spaces, the homotopy axiom holds for the Alexander-Kolmogoroff groups. Originally our proof was so devised as to apply directly to the Alexander-Kolmogoroff groups. This, however, left to one side the question of the equivalence of the Čech groups and the Alexander-Kolmogoroff groups. A positive answer was given by Spanier[8] for compact Hausdorff spaces. Our present method is this—we show that the unrestricted Čech groups are the same as the Alexander-Kolmogoroff groups. We thus slay several dragons with one stroke, obtaining Spanier's result as a corollary and getting the homotopy theorem by using Dowker's result. Looking at it from another direction we know that the extension and reduction theorems[9] hold for the unrestricted Čech groups over fully normal spaces. Hence the map excision theorem holds. Moreover we see at once that if a space has dimension at most n, then its groups in dimensions above n all vanish. Further, the groups of convex subsets of linear metric spaces all are trivial. It is thus quite plausible that one may construct an index-theory of the Leray-Schauder type for functions that are not completely continuous, since there are no compactness hypotheses in our theorems. Also it is hopeful that much of the Morse critical point theory may be developed in terms of the Alexander-Kolmogoroff groups.

Our main result is the theorem: *For fully normal spaces the Alexander-Kolmogoroff cohomology theory agrees with the unrestricted Čech cohomology theory for arbitrary coefficient groups.* A brief sketch of its proof is given as follows.

The notations of the main literature[10] will be used. In order to distinguish the Alexander-Kolmogoroff cohomology groups and the unrestricted Čech cohomology groups, we denote the latter by $\check{H}^p(X, A)$. β and γ with supplementary indices denote natural homomorphisms from cocycles to cohomology classes. Let $\sigma = (\sigma_1, \sigma_2)$ be a covering of a pair (X, A) and (K_σ, L_σ) its nerve. In the definition of $H^p(K_\sigma, L_\sigma)$ we may use[11] ordered simplexes instead of oriented simplexes. The natural homomorphism of $H^p(K_\sigma, L_\sigma)$ into $\check{H}^p(X, A)$ will be denoted by π_σ. A covering σ of (X, A) is said to be canonical if (i) $U \in \sigma_2$ implies $U \cap A \neq \phi$, (ii) $U \in \sigma_1 - \sigma_2$ implies $U - A \neq \phi$ and (iii) there is a $1 - 1$ function j_σ from the vertices of K_σ to X such that $j_\sigma(U) \in U \cap A$ or $j_\sigma(U) \in U - A$ according to $U \in \sigma_2$ or $U \in \sigma_1 - \sigma_2$. A *-refinement of a covering σ of (X, A) is a covering ρ of (X, A) such that $\rho > \sigma$, $\rho_1^* > \sigma_1$ and $(\rho_2 \wedge A)^* > \sigma_2 \wedge A$.[12] Clearly every covering of a fully normal pair (X, A) (i.e., both X and A are fully normal) has a *-refinement.

Let (X, A) be fully normal. Define a multi-valued function $\kappa: \Phi_Z{}^p(X, A) \to \check{H}^p(X, A)$ such that $\kappa\varphi$ consists of all the elements of the form $\pi_\sigma \gamma_\sigma j_\sigma{}^\# \varphi$,

where σ is a canonical covering of (X, A) such that $\varphi = 0$ on $N_{p+1}((\sigma_2 \wedge A)^{**})$ and $\delta\varphi = 0$ on $N_{p+2}(\sigma_1^{**})$.

LEMMA 1. *κ is a homomorphism with $\kappa\Phi_H^p(X, A) = 0$. Hence κ induces a homomorphism $\kappa^*: H^p(X, A) \to \check{H}^p(X, A)$ such that $\kappa = \kappa^*\gamma$.*

Clearly $\kappa\varphi$ is not empty. In order to prove the single-valuedness of κ let $\pi_\sigma\gamma_\sigma j_\sigma^\sharp\varphi, \pi_\rho\gamma_\rho j_\rho^\sharp\varphi \in \kappa\varphi$ with $\rho > \sigma$. Let $\pi_{\rho\sigma}$: $(K_\rho, L_\rho) \to (K_\sigma, L_\sigma)$ be a projection and define g: $X \to X$ such that $g(x) = x$ for $x \neq j_\rho(V)$ and $g(j_\rho(V)) = j_\sigma\pi_{\rho\sigma}(V)$. It follows[13] that $j_\rho^\sharp\varphi - j_\rho^\sharp g^\sharp\varphi$ is a coboundary. Hence $\gamma_\rho j_\rho^\sharp\varphi = \pi_{\rho\sigma}{}^*\gamma_\sigma j_\sigma^\sharp\varphi$ and $\pi_\sigma\gamma_\sigma j_\sigma^\sharp\varphi = \pi_\rho\gamma_\rho j_\rho^\sharp\varphi$.

Given $\varphi, \varphi' \in \Phi_Z^p(X, A)$ there is some canonical covering σ of (X, A) such that $\kappa\varphi = \pi_\sigma\gamma_\sigma j_\sigma^\sharp\varphi$ and $\kappa\varphi' = \pi_\sigma\gamma_\sigma j_\sigma^\sharp\varphi'$. Hence κ is a homomorphism.

If $\varphi \in \Phi_H^p(X, A)$, then there is some canonical covering σ of (X, A) such that $\kappa\varphi = \pi_\sigma\gamma_\sigma j_\sigma^\sharp\varphi$ and $\gamma_\sigma j_\sigma^\sharp\varphi = 0$. Hence $\kappa\Phi_H^p(X, A) = 0$.

LEMMA 2. *If f_μ: $(X, A) \to (Y, B)$ is a mapping, then $f^*\kappa^* = \kappa^* f^*$.*

The lemma is proved by fixing $\varphi \in \Phi_Z^p(Y, B)$ and letting $\kappa\varphi = \pi_\sigma\gamma_\sigma j_\sigma^\sharp\varphi$ for some canonical covering σ of (Y, B). There is a canonical covering α of (X, A) such that $\alpha > \sigma$ and $\kappa f^\sharp\varphi = \pi_\alpha\gamma_\alpha j_\alpha^\sharp f^\sharp\varphi$. Define $f_{\alpha\sigma}$: $(K_\alpha, L_\alpha) \to (K_\sigma, L_\sigma)$ such that $f_{\alpha\sigma}(U) = V$ implies $f(U) \subset V$. Define g: $X \to Y$ such that $g(x) = f(x)$ for $x \neq j_\alpha(U)$ and $g(j_\alpha(U)) = j_\sigma f_{\alpha\sigma}(U)$. As in Lemma 1 we have $\gamma_\alpha j_\alpha^\sharp g^\sharp\varphi = \gamma_\alpha j_\alpha^\sharp f^\sharp\varphi$ and hence $\pi_\alpha\gamma_\alpha j_\alpha^\sharp f^\sharp\varphi = f^*\pi_\sigma\gamma_\sigma j_\sigma^\sharp\varphi$, or $\kappa^* f^*(\gamma\varphi) = f^*\kappa^*(\gamma\varphi)$.

LEMMA 3. $\delta\kappa^* = \kappa^*\delta$.

We have only to consider the case $X \neq A$. Let i: $A \to X$ be the injection. For a given $\varphi \in i^{\sharp -1}\Phi_Z^p(A)$ there is a collection α_2 of open sets in X such that $\kappa i^\sharp\varphi = \pi_\sigma\gamma_\sigma j_\sigma^\sharp i^\sharp\varphi$ with $\sigma_1 = \alpha_2 \wedge A$. Let $\alpha_1 = \alpha_2 \cup \{X\}$ and define j_α such that $j_\alpha(U) = j_\sigma(U \cap A)$, $U \in \alpha_2$, and $j_\alpha(X) \in X - A$. Then $\kappa\delta^\flat\varphi = \pi_\alpha\gamma_\alpha j_\alpha^\sharp\delta^\flat\varphi$. It follows from the definition of the coboundary operator and the various permutability conditions that $\pi_\alpha\gamma_\alpha j_\alpha^\sharp\delta^\flat\varphi = \delta\pi_\sigma\gamma_\sigma j_\sigma^\sharp i^\sharp\varphi$, or $\kappa^*\delta(\beta i^\sharp\varphi) = \delta\kappa^*(\beta i^\sharp\varphi)$.

LEMMA 4. κ^*: $H^p(X) \approx \check{H}^p(X)$.

The idea of the proof is as follows. Given $\varphi \in \Phi_Z^p(X)$ with $\kappa\varphi = 0$ there is some canonical covering σ of (X, A) such that $\kappa\varphi = \pi_\sigma\gamma_\sigma j_\sigma^\sharp\varphi$ and $\gamma_\sigma j_\sigma^\sharp\varphi = 0$. Let $\rho_1{}^* > \sigma_1$ and define k_σ from X to the vertices of K_σ such that $k_\sigma(x) \in U$ implies $x \in V \subset V^* \subset U$ for some $V \in \rho_1$. Hence[13] $\gamma\varphi = \gamma(j_\sigma k_\sigma)^\sharp\varphi = 0$. Given any $\pi_\sigma\gamma_\sigma\varphi_\sigma \in \check{H}^p(X)$ let $\rho_1{}^* > \sigma_1$ and defined k_σ as above. Then $\varphi = k_\sigma^\sharp\varphi_\sigma \in \Phi_Z^p(X)$ and $\kappa\varphi = \pi_\tau\gamma_\tau j_\tau^\sharp\varphi = \pi_\sigma\gamma_\sigma\varphi_\sigma$, where τ is a canonical covering of (X, ϕ) with $\tau_1{}^{**} > \rho_1$.

LEMMA 5. κ^*: $H^p(X, A) \approx \check{H}^p(X, A)$.

The lemma is an automatic consequence of the preceding three lemmas and the exactness of cohomology sequences.

Combining Lemmas 2, 3 and 5, our theorem is proved.

After getting this result I heard that Professor Dowker obtained an analogous result.

* This work was done under contract N7-onr 434 Task Order III, Naval Department, Office of Naval Research.

[1] Cartan, H., "Algebraic Topology," Harvard Univ., 1949; Wallace, A. D. "Outline for Algebraic Topology I," Tulane Univ., 1949–1950; and Wallace A. D., "Map Excision Theorem," *Duke Math. J.*, 19 (1952).

[2] Cf. Tukey, J. W., "Convergence and Uniformity in General Topology," *Ann. Math. Studies*, Princeton, 1940, p. 53.

[3] Cf. Tukey, J. W., *loc. cit.*

[4] See Stone, A. H., "Paracompactness and Product Spaces," *Bull. Am. Math. Soc.*, **54,** 977–982 (1948).

[5] Cf. Spanier, E. H., "Cohomology Theory for General Spaces," *Ann. Math.* (2), **49,** 407–427 (1948).

[6] Cf. Wallace, A. D., *loc. cit.*

[7] See Dowker, C. H., "Čech Cohomology Theory and the Axioms," *Ibid.* (2) **51,** 278–292 (1950).

[8] Cf. Spanier, E. H., *loc. cit.*

[9] Cf. Wallace, A. D., *loc. cit.*

[10] Spanier, E. H., *loc. cit.* and Dowker, C. H., *loc. cit.* (2) **51,** 418.

[11] Cf. Eilenberg, S., "Singular Homology Theory," *Ibid.*, (1944), Cor. 9.3.

[12] $\rho_2 \wedge A = \{U \cap A \mid U \in \rho_2\}$.

[13] Make use of a homotopy lemma similar to Spanier, *loc. cit.*, Lemma 9.1.

ON THE EARLY CHEMICAL HISTORY OF THE EARTH AND THE ORIGIN OF LIFE

BY HAROLD C. UREY

INSTITUTE FOR NUCLEAR STUDIES, UNIVERSITY OF CHICAGO

Communicated January 26, 1952

In the course of an extended study on the origin of the planets[1] I have come to certain definite conclusions relative to the early chemical conditions on the earth and their bearing on the origin of life. Oparin[2] has presented the arguments for the origin of life under anaerobic conditions which seem to me to be very convincing, but in a recent paper Garrison, Morrison, Hamilton, Benson and Calvin,[3] while referring to Oparin, completely ignore his arguments and describe experiments for the reduction of carbon dioxide by 40 m. e. v. helium particles from the Berkeley 60-inch cyclotron. As I believe these experiments, as well as many previous ones using ultra-violet light to reduce carbon dioxide and water and giving similar results to theirs, are quite irrelevant to the problem of the origin of life, I wish to present my views.

During the past years a number of discussions on the spontaneous origin of life have appeared in addition to that by Oparin. One of the most extensive and also the most exact from the standpoint of physical chemistry

is that by Blum.[3a] It seems to me that his discussion meets its greatest difficulty in accounting for organic compounds from inorganic sources. This problem practically disappears if Oparin's assumptions in regard to the early reducing character of the atmosphere are adopted.

In order to estimate the early conditions of the earth, it is necessary to ask and answer the questions of how the earth originated, and how the primitive earth developed into the present earth. The common assumption is that the earth and its atmosphere have always been as they are now, but if this is assumed it is necessary to account for the present highly oxidized condition by some processes taking place early in the earth's history. Briefly, the highly oxidized condition is rare in the cosmos and exists in the surface regions of the earth and probably only in the surface regions of Venus and Mars. Beyond these we know of no highly oxidized regions at all, though undoubtedly other localized regions of this kind exist. This is essentially the argument of Oparin.

The surface of the moon gives us the most direct evidence relative to the origin of the earth. Gilbert[4] called attention to the great system of ridges and grooves radiating from Mare Imbrium, and Baldwin[5] has explained this as due to a colliding planetesimal some hundred kilometers in radius. My own studies show that the object contained metallic iron objects and and silicate materials and that water as such or as hydrated silicates arrived with such objects. This collision occurred during the terminal stage of the moon's formation. Some five other similar objects left their marks on the moon's surface and they all fell within a time span of some 10^5 years. Their temperatures were not high; probably not appreciably higher than present terrestrial temperatures. The arguments for these conclusions are long and detailed and cannot be repeated here.

At the time such objects were falling on the moon similar objects fell on the earth. The conditions were different because of the greater energy of such objects, 22 times as great if they fell from a great distance and 11 times as great if they fell from the circum surface orbit, and because of the presence of a substantial atmosphere on the earth. The energy was sufficient to completely volatilize the colliding planetesimals and raise the gas to greater than 10,000°K. An object similar to the Imbrium planetesimal would have distributed its material over a region several thousand kilometers in linear dimensions and the explosion cloud would have risen far above the atmosphere. Its materials would have fallen through the atmosphere in the form of iron and silicate rains and would have reacted with the atmosphere in the process. (H. H. Nininger recently showed me spherical iron-nickel objects collected near Meteor Crater, Ariz., which were formed in such a rain.) The objects contained metallic iron-nickel alloy, silicates, graphite, iron carbide, water or water of crystallization, ammonium salts and nitrides, that is substances which would supply the volatile and non-

volatile constituents of the earth. The temperatures produced in these collisions were very high, but unless the accumulation was very rapid indeed the general temperature of the planet was not excessively high. That such objects fell on the moon and earth at the terminal stage of their formation I regard as certain, and it is difficult in this subject to be certain about anything. But regardless of the detailed arguments, those who postulate oxidizing conditions as the initial state of the earth should present some similar argument to justify their assumption.

The reactions taking place at that time of interest to us here are:

$FeO + H_2 = Fe + H_2O$;	$K_{298} = 1.7 \times 10^{-3}$, $K_{1200} = 0.97$
$Fe_3O_4 + H_2 = 3FeO + H_2O$;	$K_{298} = 2.5 \times 10^{-2}$, $K_{1200} = 1.0$
$C + H_2O = CO + H_2$;	$K_{298} = 10^{-16}$, $K_{1200} = 3.8 \times 10^{-6}$
$C + 2H_2 = CH_4$;	$K_{298} = 7.8 \times 10^{8}$, $K_{1200} = 1.6 \times 10^{-2}$
$Fe_3C + 2H_2 = Fe + CH_4$;	$K_{298} = 3.2 \times 10^{11}$, $K_{1200} = 5.9 \times 10^{-3}$
$NH_3 = {}^1/_2N_2 + {}^3/_2H_2$;	$K_{289} = 1.2 \times 10^{-3}$.

From these equilibrium constants one sees that hydrogen was a prominent constituent of the primitive atmosphere and hence that methane was as well. Nitrogen was present as nitrogen gas at high temperatures but may have been present as ammonia or ammonium salts at low temperatures. At high temperatures hydrogen would escape from the planet very rapidly, but if the temperatures were high objects must have arrived rapidly to replenish the lost hydrogen. If the objects arrived slowly then the temperatures were low and hydrogen did not escape rapidly. Thus it is very difficult to see how the primitive atmosphere of the earth contained more than trace amounts of other compounds of carbon, nitrogen, oxygen and hydrogen than CH_4, H_2O, NH_3 (or N_2) and H_2.

We now consider what could reasonably be expected to convert this atmosphere into the present one existing on the earth.* If there was a large

* Poole, J. H. J. (*Proc. Roy. Soc. Dublin*, **22**, 345 (1941)), Harteck, P., and Jensen, J. H. D. (*Z. Naturforschung*, **3a**, 581 (1948)), and Dole, M. (*Science*, **109**, 77 (1949)) have reviewed the modern evidence for the photochemical origin of free oxygen and the evidence against the photosynthetic origin. Their conclusions are accepted here. In a recent paper, Poole, J. H. J. (*Sci. Proc. Roy. Dublin Acad.*, **25**, 201 (1951)) definitely concludes that methane and ammonia were not present in the primitive atmosphere and that it consisted of H_2O, CO_2 and N_2 but contained no oxygen. I cannot accept this conclusion for carbon dioxide and nitrogen are almost as difficult to understand as free oxygen. The interior of the earth and the lavas which reach its surface are highly reducing and are not a likely source of highly oxidized materials. Poole assumes that carbon dioxide was present in large quantities in the primitive atmosphere. How was it produced from methane, graphite or iron carbide? He also assumes that it would remain in the atmosphere, but as is shown in the text and the references cited, carbon dioxide in the presence of water should react rapidly with silicates until its partial pressure reaches values in the neighborhood of those on the earth and Mars.

amount of hydrogen, the outer parts of the atmosphere beyond the convection zone would become highly enriched in hydrogen. Hydrogen would absorb light from the sun in the far ultra-violet[6] and since it does not radiate in the infra-red would become a high-temperature atmosphere just as exists on the earth at present, and hydrogen would be lost very rapidly. As hydrogen was depleted, the atmosphere would become a methane one and since methane and its photochemical disintegration products absorb a wide band of energy in the ultra-violet and radiate in the infra-red, the temperature of the high atmosphere would fall far below the present temperature of 1500° or 2000°C.[7] In fact, it might well approach present terrestrial surface or even lower temperatures. The loss of hydrogen would be decreased, but since an oxidized atmosphere is present on the earth it can be assumed that it escaped at some appropriate rate. The net process was the dissociation of water into hydrogen which escaped, and into oxygen which oxidized reduced carbon compounds to carbon dioxide, ammonia to nitrogen, and reduced iron to more oxidized states. When the methane and ammonia were oxidized free oxygen appeared and the present atmospheric conditions were established. As carbon dioxide was formed it reacted with silicates to form limestones, i.e.,

$$CaSiO_3 + CO_2 = CaCO_3 + SiO_2, \qquad K_{298} = 10^8.$$

Of course the silicates may have been a variety of minerals but the pressure of CO_2 was always kept at a low level by this reaction or similar reactions just as it is now. Plutonic activities reverse the reaction from time to time, but on the average the reaction probably proceeds to the right as carbon compounds come from the earth's interior,[8] and in fact no evidence for the deposition of calcium silicate in sediments seems to exist.†

The histories of Mars and Venus should be similar to that of the earth. Mars has no mountains higher than 750 meters. Thus the initial lunar type mountains were probably eroded by water and no folded mountains have been formed. The oxidation of methane to carbon dioxide and the

† The alternative to this course of events would be the production of carbon dioxide from the earth's interior and there is evidence that at least some oxidized carbon is so produced. Gases escaping from lava lakes of Hawaiian volcanoes are highly oxidized, so much so that it is difficult to account for the high states of oxidation (Day, A. L., and Shepherd, E. S., *J. Wash. Acad. Sci.*, **3**, 457 (1913); Shepherd, E. S., *Bull. Hawaiian Volcano Obs.*, **VII**, 97 (1919); *Ibid.*, **VIII**, 65 (1920).) It seems probable to the writer that atmospheric oxygen directly or indirectly is responsible for the oxidation. Observers report burning gases as escaping from lavas, thus indicating the escape of reduced gases from the lava. Also surface oxidation of the lava pool must occur, as is shown by the high temperatures in the surfaces of such pools. Ferric oxide is probably produced and this in turn would oxidize carbon within the magma. (See *Bull. Hawaiian Volcano Obs.*, **IX**, 113 (1921).) If atmospheric oxygen is the source, then the carbon dioxide does not represent oxidized carbon being supplied to the atmosphere, but on the contrary it represents reduced carbon being so supplied.

formation of limestone proceeded as on earth, but oxygen atoms or water molecules were lost from the planet. In an atmosphere containing oxygen and nitrogen, a high temperature should have existed on Mars just as exists on the earth, and due to the smaller gravitational field atoms of atomic weight, 16, should escape if atoms of atomic weight, 4, i.e. He, escape from the earth now, as they do Finally, there results a desert planet with very small amounts of water and a pressure of carbon dioxide in its atmosphere about equal to that on the earth,[9] the excess carbon dioxide having reacted with the silicates to form limestones until its partial pressure was reduced to a low value. Since mountains are absent, volcanic activity must be small or non-existent and carbon dioxide has not been generated from limestone and silicon dioxide.

Venus started with a reduced atmosphere, which was oxidized to carbon dioxide and limestone by photochemical action. It cannot lose water, however, and the absence of water means that much less water was present initially than in the case of the earth, probably due to having been formed nearer to the sun and thus at a higher temperature. Assuming plutonic activity on Venus, the carbon dioxide has been regenerated by processes similar to those of the earth. In the absence of water the reaction of carbon dioxide with silicates is very slow and hence a dense atmosphere of carbon dioxide is possible. Thus a reasonable course of events can be postulated for this planet.‡

The Origin of Life.—The problem of the origin of life involves three separate questions in our present discussion: (1) the spontaneous formation of the chemical compounds which form the physical bodies of living organisms; (2) the evolution of the complex chemical reactions which are the dynamic basis of life; and (3) the source of free energy which alone can maintain the chemical reactions and synthesize the chemical compounds. It is only the first and third questions which will be discussed here. At present the source of free energy is sunlight through photosynthesis, but how were primitive living organisms maintained through a long enough period of time for the evolution of photosynthesis to occur?

It is suggested here that life evolved during the period of oxidation of highly reduced compounds to the highly oxidized ones of today. During

‡ The discussion up to this point, together with the suggestion that life originated during the period of oxidation of reduced carbon compounds to oxidized ones, was presented before the Geological Society of America, Washington, November, 1950. It was thought that there might be some condition of pressure, temperature and composition such that organic compounds became stable, thus making the synthesis of complex compounds possible. The general ideas were discussed with Dr. H. E. Suess, who was a Fellow at the Institute for Nuclear Studies. He made some studies relative to this problem which did not appear to be promising of positive results. The present paper follows a somewhat different approach to the problem by assuming the synthesis of organic compounds by means of ultra-violet light in the high atmosphere.

this period compounds of carbon, oxygen, hydrogen, and nitrogen were present in substantial amounts. It is also suggested that the source of free energy was the absorption of ultra-violet light in the high atmosphere by methane and water and other compounds produced from them by this photochemical action. This process also protected primitive organisms from ultra-violet light in the absence of an ozone layer as exists now.

Organic compounds are generally unstable relative to completely reduced or completely oxidized compounds throughout the entire range of hydrogen and oxygen pressures in chemical equilibrium with water at ordinary temperatures, but photochemical processes should produce such compounds. The exact conditions obtaining in a methane atmosphere can only be roughly estimated. Convection to higher altitudes than now occur on the earth should have been present, since a methane atmosphere radiates in the infra-red while an oxygen-nitrogen atmosphere does not. The atmosphere should have been cooler at high altitudes so that convection extended to higher levels than now, and water vapor should have been carried to high altitudes and photochemical products of the high atmosphere should have been moved rapidly downward.

The photochemical processes in a pure methane atmosphere can be estimated qualitatively. Methane absorbs in the ultra-violet below 1450 A. The total energy of the present solar spectrum below this wave-length is about 5×10^{-5} of the total energy. Methane dissociates into methyl and atomic hydrogen. Methyl probably absorbs at much longer wave-lengths and probably repulsive states exist resulting in the formation of methylene. This compound likewise would be dissociated into CH. Thus the reactions

$$\begin{aligned} CH_4 + h\nu\ (\lambda < 1500) &= CH_3 + H \\ CH_3 + h\nu\ (\lambda < 2800) &= CH_2 + H \\ CH_2 + h\nu\ (\lambda < 2800) &= CH + H \\ CH + h\nu\ (\lambda < 2800) &= C + H \end{aligned}$$

will occur and with their reversal and the reaction of the primary products with each other and with secondary products a steady state of great complexity will be established, the details of which cannot be estimated because of many unknown factors. The absorption spectra of CH_3 and CH_2 and the kinetics of the back reactions and other reactions are unknown. The fraction of the sun's energy below 2000 A is 3.3×10^{-4} and below 2500 A is 2.2×10^{-3}, so that very appreciable dissociation of CH_3 and CH_2 may be expected. The energies per year cm.$^{-2}$ of the earth's surface for complete absorption of the sun's spectrum at the earth below 1500 A, 2000 A and 2500 A are:

	Cal. $yr.^{-1}$ $cm.^{-2}$
1500	1.6
2000	85
2500	570

Not much methyl would be produced directly but secondary reactions such as $CH + CH_4 = CH_3 + CH_2$ could be expected to produce more of this radical. Altogether a very considerable absorption of solar energy would be expected.

In a pure water atmosphere dissociation of water into hydroxyl and atomic hydrogen would occur by absorption of light below 1900 A. Also dissociation of hydroxyl should occur giving atomic oxygen. Secondary reactions should produce O_2, H_2O_2, HO_2 and O_3 in amounts very difficult to estimate.

A combined methane and water atmosphere quite obviously would give a great variety of compounds of carbon, hydrogen and oxygen. In particular the reaction

$$CH_4 + OH = CH_3 + H_2O$$

would occur, thus producing larger quantities of methyl. These compounds would move by convection to lower levels in some quantities, dissolve in rain water and produce solutions of organic compounds in the oceans. Here ammonium salts should have been present (see below), and and the formation of nitrogen-containing compounds would occur. Given time, some natural catalysts and very slow destruction of organic compounds because of the absence of living organisms, a large number of organic compounds would be expected. If all the present surface carbon were dissolved in the present oceans as organic compounds, the oceans would become approximately a one per cent solution of these compounds. Thus compounds suitable for living organisms were possible and probably abundant.

Though the conditions postulated above are not approximated in any past experiments so far as I have been able to determine, the extensive studies on photochemistry, free radicals produced by various methods and the effects of electrical discharges on chemical substances[10] leave no doubt that many compounds would be formed due to the absorption of ultraviolet light.

A free energy supply for primitive living organisms is necessary, for only in this way can an active metabolism be supported and in the absence of such metabolism only dead and not living organic substance is possible. Rabinowitch[11] estimates that the present annual energy from photosynthesis is 600 cal. $cm.^{-2}$ of the earth's surface. The following table shows the standard free energies for three types of reactions of carbon-hydrogen-oxygen compounds of different oxidation states of the carbon atom.

COMPOUND	DISPROPORTION TO CH_4, CO_2 AND H_2O	OXIDATION TO CO_2 AND H_2O	REDUCTION TO CH_4 AND H_2O
CH_4	0	−195.50	0
CH_3OH	−22.31	−168.94	−30.13
CH_2O	−27.0	−124.75	−42.63
CH_2O_2	−21.83	−70.71	−45.28
CO_2	0	0	−31.26
$^1/_6C_6H_{12}O_6$	−17.23	−114.97	−32.85

It is apparent from this table that substantial quantities of free energy are available from the disproportion and reduction reactions of organic compounds, i.e., reactions possible under reducing conditions, though the free energies of the oxidation reactions possible in the present oxidizing atmosphere are much larger. It is not intended to infer that the reactions listed are necessarily the ones used by primitive life, but only that they indicate the order of magnitude of the free energies available from similar reactions. (Yeast uses a disproportion reaction,

$$^1/_6C_6H_{12}O_6(aq.) = {}^1/_3C_2H_5OH + {}^1/_3CO_2, \qquad \Delta F^\circ_{298} = -8.68$$

as a source of its free energy.) The high energy photochemical reactions of the reducing atmosphere at high altitudes could not be highly effective because of back reactions and because of the high energy used for the elementary processes (~150,000 cal.). Hence, the free energy supply for the primitive life processes was much less than that of the present time probably not more than 10^{-3} or 10^{-4} as much. However, experimentation on metabolic processes was possible and probably proceeded on a substantial scale. Also, a great advantage accrued to the mutations producing photosynthesis, thus ensuring survival of these processes.

Porphyrins probably appeared during the reducing period as important constituents of enzymes. Also during this time, photosynthesis evolved, and as oxidizing conditions were established green plants became the fundamental, even if they are not the dominant, type of life. In this way the evolution of photosynthesis was possible before the free energy due to it was available to living organisms.

Time and Conditions of Transition Period.—The order in which reduced substances were oxidized depends on the free energies of the oxidation reactions,

$$^1/_2CH_4(g.) + O_2 = {}^1/_2CO_2 + H_2O, \qquad \Delta F^\circ_{298}\ \text{kcal.} = -97.75$$
$$^4/_3NH_3(g.) + O_2 = {}^2/_3N_2 + 2H_2O, \qquad \Delta F^\circ_{298}\ \text{kcal.} = -110.73$$
$$^1/_2H_2S + O_2 = {}^1/_2H_2SO_4\ (1\ \text{molar}), \qquad \Delta F^\circ_{298}\ \text{kcal.} = -84.72$$
$$^1/_2FeS + O_2 = {}^1/_2\ FeSO_4\ (aq.), \qquad \Delta F^\circ_{298}\ \text{kcal.} = -86.7.$$

Thus ammonia should oxidize first, then methane, and hydrogen sulfide third. However, if ammonium salts of organic acids are possible the stability of reduced nitrogen is greatly increased, and thus methane and ammonium ion

would be oxidized more or less together. This is shown by the reaction,

$$^4/_3NH_4OOCCH_3+O_2 = {}^2/_3N_2+2H_2O+{}^4/_3CH_3COOH, \ \Delta F^\circ_{298} = -97.83.$$

The last two reactions show that sulfide sulfur will be oxidized after the methane and ammonium ion have been oxidized. This conclusion is supported by other oxidation-reduction reactions of sulfur and carbon.

The amount of carbon on the earth's surface is about 350 g. atoms cm.$^{-2}$ equivalent to 1.4×10^3 moles cm.$^{-2}$ or 2.5×10^4 g. cm.$^{-2}$ of water, if all carbon was initially present as CH_4 and all had been oxidized to CO_2. But carbon is produced from the earth's interior and its state of reduction is less than that of methane and part of the surface carbon is not now oxidized to carbon dioxide. Hence the total amount of water which has been decomposed in order to oxidize the carbon is more nearly half of the above value, or 700 moles cm.$^{-2}$ of water. Using the energy per year absorbed by the atmosphere in wave-lengths below 1500 A, i.e., 1.6 cal. yr.$^{-1}$ cm.$^{-2}$, and assuming that every quantum produces a hydrogen atom with 200,000 cal. per/gram atom and that the hydrogen atom escaped from the earth, 10^8 years would be required for the oxidation process. If effectively all the energy below 2000 A is utilized in this way, the time would be 2.5×10^6 years. Neither assumption is realistic and no estimate can be made in this way except that the time might be either very long or comparatively short.

The hydrogen must escape in order for an oxidized atmosphere to be established, and if the methane-water atmosphere is cold, escape will be difficult. Interpolating from Spitzer's calculations[12] the escape of the required amount of hydrogen would require about 2×10^9 years, if atomic hydrogen escaped at 325°K. with an effective surface partial pressure of 10^{-3} atmosphere and 2×10^6 years if the surface pressure was one atmosphere. In the latter case the escape formula is not a good approximation but the time would be short nevertheless. If escape was by molecular hydrogen the temperature must be 650°K. for the same times of escape. The temperature of the methane water atmosphere at high altitudes was probably less than 325°K. and hence a long time for the escape of hydrogen from the reducing atmosphere is indicated.

Thode[13] and his coworkers have found that the ratios of the sulfur isotopes in the sulfides, elementary sulfur and sulfates are closely the same as this ratio in meteoritic sulfur until about 8×10^8 years ago, and after this time the sulfur and sulfides contain increasing amounts of S^{32} with time while the sulfates contain less amounts of this isotope. They ascribe this to the action of living organisms in promoting the oxidation and reduction of sulfur compounds, thus leading to a progressive separation of the isotopes, and suggest that life evolved about 8×10^8 years ago. This is a very interesting suggestion and may be a correct conclusion. It does

not give a very long time for the evolution of the very comlpex organisms whose remains are found in the Cambrian rocks.

On the other hand, this date might mark the transition from the reducing to the oxidizing atmospheric conditions. The oxidation of sulfur and sulfides to sulfates would probably not occur to any large extent until free oxygen appeared or until photosynthesis was well developed. But limestones were deposited in large quantities early in the earth's history and graphite was not. The two reactions,

$$CaCO_3 + SiO_2 + 4H_2(g.) = CH_4 + 2H_2O + CaSiO_3, \qquad K = 3 \times 10^{16}$$

and

$$C + 2H_2(g.) = CH_4(g.), \qquad K = 8 \times 10^8,$$

make possible an estimate of the pressure of hydrogen and methane that would make these two events possible at the same time. The pressure of methane would only be 5 atmospheres if all the present surface carbon were methane and if part of this carbon were dissolved as organic compounds in the oceans, the partial pressure might well be about one atmosphere. Then, if limestones were deposited, the hydrogen partial pressure was less than 10^{-4} atmosphere if equilibrium existed, but was probably higher since complete equilibrium cannot be expected. The second equation shows that graphite would not be stable under these conditions. As the methane was consumed the hydrogen pressure must have decreased. The partial pressure of carbon dioxide was 10^{-8} atmosphere if calcium carbonate was present, and may have been higher than this as it is today. If the hydrogen pressure fluctuated and for brief periods exceeded some critical pressure, massive deposits of limestones would be possible, but organisms which experimented with calcareous shells would have had great difficulty in preventing the dissolution of their shells during these periods and the extinction of their species, and indeed no certain calcareous fossils have been found in the Precambrian.

It seems just barely possible that reducing conditions were maintained until some 8×10^8 years ago. Limestones could be deposited, graphite need not have been formed, living organisms might find some 10^{-3} atmosphere or even less of hydrogen with photochemically oxidized organic compounds sufficient for their metabolic processes, and the methane pressure could not have been above 5 atmospheres and was maintained at lower pressures by the solubility of the oxidized organic compounds in water. The precipitation of limestones in great quantities presents a difficulty to the hypothesis of a long period during which a reducing atmosphere was present. The presence of highly oxidized iron in "red beds" and hematite (Fe_2O_3) iron ores are justly regarded as evidence for atmospheric oxygen. Red beds apparently are unknown earlier than the late Precambrian.

Most of the great bodies of iron ore were laid down in the late Precambrian (Huronian) or were extensively eroded during this time. The iron ore of the Vermillion range of Minnesota is much earlier (Keewatin) and thus oxidation of ferrous iron to ferric oxide took place early in the earth's history.[14] It should be noted that ferrous oxide should be oxidized to magnetic iron oxide by water if the temperature is sufficiently high to make the reaction fast enough with respect to the time available. However, magnetic iron oxide cannot be oxidized to ferric oxide by water unless the hydrogen is removed. The relations are shown by the reactions,

$$3FeO + H_2O = Fe_3O_4 + H_2; \qquad K_{298} = 10^8, \quad K_{500} = 10^4$$
$$2Fe_3O_4 + H_2O = 3Fe_2O_3 + H_2; \qquad K_{298} = 10^{-6}, \quad K_{500} = 10^{-5}$$

Thus circulating hot water could produce ferric oxide even in the absence of free oxygen, but it would probably be a rare event. The magnetic iron oxide is formed when iron is corroded by water in boilers. The conditions of deposition of these ores of the Precambrian are not well understood, though as stated above the presence of highly oxidized iron justifies a strong presumption of an oxidizing atmosphere. Thoughout the calculations it has been assumed that thermodynamic equilibria will be attained except for photochemical effects. This need not be the case and the presence of living organisms almost certainly would lead to important deviations from such equilibrium.

The red bacteria and some species of algae are able to use hydrogen and carbon dioxide in photosynthesis. This ability to use hydrogen is especially interesting because they do not find hydrogen available in their natural habitats. They appear to be living fossils from some former time and would live under conditions outlined above, though they prefer higher pressures of hydrogen than 10^{-3} atmosphere. Incidentally, modern plants prefer higher concentrations of carbon dioxide than those available in nature.§ If the present atmosphere should slowly change to a reducing one, it is certain that a substantial flora and fauna would survive. The flora would surely include many green plants and the fauna most of the principal orders of animals with the exception of the mammals and birds, i.e., the warm-blooded animals, for whom the reduced free energy supply would probably be fatal. A few aerobes would probably survive wherever photosynthesis was very intense. Aerobic organisms must naturally be most abundant under aerobic conditions, but mutations would surely supply anaerobic ones for life in a reducing atmosphere.[15]

Poole[16] thinks that oxygen may have been absent from the earth's atmosphere for some 10^9 years of the earth's history, but according to the evidence given here his model of the primitive atmosphere is not correct and hence his conclusion does not substantiate the present work. He shows that Tammann's thermal dissociation is not correct and that photo-

chemical dissociation and the escape of hydrogen are necessary for the formation of free oxygen. It is contended here that this mechanism is necessary to account for carbon dioxide as well. Lane[17] has argued that free oxygen did not appear until late Precambrian times because of the reduced condition of the Keewatin Greenstone schists. MacGregor[18] comes to similar conclusions from the Precambrian rocks of Rhodesia. He suggests that the Precambrian iron deposits were concentrated from igneous surface rocks by solution of iron in the absence of free oxygen as ferrous carbonate which was precipitated as ferric oxide by the action of green plants in a lake or sea into which the rivers ran. The plants may have been diatoms and hence the well-known banded structure of hematite and jasper may have been produced. As indicated above, the origin of these deposits is not well understood and therefore these suggestions, while worthy of consideration, cannot be regarded as conclusive.

The general course of events and the favorable condition for the origin of life outlined in this paper in no way depend on the time of transition from reducing to oxidizing conditions being exactly some 8×10^8 years ago. However, the evolution from inanimate systems of biochemical compounds, e.g., the proteins, carbohydrates, enzymes and many others, of the intricate systems of reactions characteristic of living organisms, and of the truly remarkable ability of molecules to reproduce themselves seems to those most expert in the field to be almost impossible. Thus a time from the beginning to photosynthesis of two billion years may help many to accept the hypothesis of the spontaneous generation of life. On the other hand, our judgment of an approximate time for the origin of life certainly is not so precise that we can say that 2×10^9 years are sufficient but 2×10^8 years are not.

It seems to me that experimentation on the production of organic compounds from water and methane in the presence of ultra-violet light of approximately the spectral distribution estimated for sunlight would be most profitable. The investigation of possible effects of electric discharges on the reactions should also be tried since electric storms in the reducing atmosphere can be postulated reasonably.

Also theoretical investigations on hydrogen and methane-water atmospheres would be most helpful in estimating the time of transition from the reducing to the oxidizing atmosphere. Most interesting in this connection would be more experimental data such as those of Dr. Thode and his coworkers on the abundance of the sulfur isotopes. The time of transition should be recorded in the rocks, and some such indication as that observed, by Thode, or some change in the state of oxidation of some elements should occur and should be detectable providing the time of transition was not too early in the earth's history.

I have profited greatly from discussions of this subject with Professor

James Franck, who has often pointed out to me and others that complex organic compounds, even porphyrins, may have originated under approximately the conditions outlined in this paper.

Note added in proof: Since this paper was written an interesting paper by J. D. Bernal on the Physical Basis of Life (London, Routledge and Kegan Paul, 1951) has come to my attention, in which very similar suggestions have been made, but there are differences in details and the arguments used. His paper is worthy of serious study.

§ I am indebted to Dr. H. Gaffron for information relative to these bacteria and algae.

[1] Urey, H. C., *Geochim. et Cosmochim. Acta*, **1**, 209 (1951). This has been revised and extended in *The Planets: Their Origin and Development*, Yale University Press, 1952.

[2] Oparin, A. I., *Origin of Life*, Macmillan Co., New York, 1938.

[3] Garrison, W. M., Morrison, D. C., Hamilton, J. G., Benson, A. A. and Calvin, M., *Science*, **114**, 416 (1951).

[3a] Blum, H. F., *Time's Arrow and Evolution*, Princeton University Press, 1951.

[4] Gilbert, G. K., *Bull. Phil. Soc. Wash.*, **12**, 241 (1893).

[5] Baldwin, R. B., *The Face of the Moon*, University of Chicago Press, 1949.

[6] See Greenstein, J. L., *Atmospheres of the Earth and Planets*, edited by G. P. Kuiper, University of Chicago Press, 1949, pp. 117 ff.

[7] See Spitzer, L., Jr., *Ibid.*, edited by G. P. Kuiper, University of Chicago Press, 1949.

[8] Rubey, W. W., *Bull. Geol. Soc. Am.*, **62**, 1111 (1951).

[9] Kuiper, G. P., *Atmospheres of the Earth and Planets*, University of Chicago Press, 1949, p. 304.

[10] See, for example, Rice, F. O., and Rice, K. K., *The Aliphatic Free Radicals*, The Johns Hopkins Press, 1935; *Free Radical Mechanisms*, Steacie, E. W. R., Reinhold, 1946; Noyes, W. A., Jr., and Leighton, P. A., *The Photochemistry of Gases*, Reinhold, 1941; Rollefson, G. K., and Burton, M., *Photochemistry*, Prentice-Hall, 1939; Glockle G., and Lind. S. C., *Electrochemistry of Gases and Other Dielectrics*, John Wiley, 1939.

[11] Rabinowitch, E. I., *Photosynthesis and Related Processes*, Vol. I, New York, 1945.

[12] Spitzer, Lyman, Jr., *Atmospheres of the Earth and Planets*, edited by G. P. Kuiper, University of Chicago Press, 1949.

[13] Thode, H. G., Am. Chem. Soc. Meeting, New York, September, 1951. See Tudge, A. P., and Thode, H. G., *Can. J. Res.*, **28**, 567 (1950).

[14] Leigh, C. K., Lund, R. J., and Leigh, Andrew, *Geol. Survey Professional Paper* 184 (1935).

[15] See Deevey, E. S., Jr., *Scientific American*, **185**, No. 4, 68 (1951) for an interesting mention of Chironomius and of other anaerobic animals.

[16] *Loc. cit.*

[17] Lane, A. C., *Am. J. Sci.*, **43**, 42 (1917).

[18] MacGregor, A., *South African J. Sci.*, **24**, 155 (1927).

SOME COMMENTS ON THE LONDON-BROOKS TREATMENT OF DIAMAGNETIC ANISOTROPY*

By G. W. Wheland and Sheldon L. Matlow

Department of Chemistry, University of Chicago

Communicated by R. S. Mulliken, January 11, 1952

In the LCAO (linear combination of atomic orbitals) approximation, a molecular orbital ψ for a system in field-free space is expressed in the form

$$\psi = \sum_{k=1}^{n} c_k \phi_k \tag{1}$$

where the constant coefficients c_k are variation parameters, and the functions ϕ_k are atomic orbitals. London[1] has shown that, when the system is in a magnetic field characterized by the vector potential $\mathbf{A}$, equation (1) can conveniently be replaced by

$$\psi = \sum_{k=1}^{n} c_k e^{\frac{2\pi e i}{hc}(\mathbf{A}_k \cdot \mathbf{r})} \phi_k \tag{2}$$

where e is the electronic charge, h is Planck's constant, and c is the velocity of light. Moreover, $\mathbf{r}$ is the radius vector with respect to any arbitrary origin, and $\mathbf{A}_k$ is the value of $\mathbf{A}$ at the point $\mathbf{r}_k$ (i.e., at the nucleus about which the orbital ϕ_k is centered). For a particular case in which each ϕ_k is a (real) p_z orbital, and in which the magnetic field, of strength H, is parallel to the z-axis, London has shown[1] also that the general element $[\mathcal{H} - W]_{kl}$ of the secular determinant can be approximated by the expression

$$[\mathcal{H} - W]_{kl} \equiv \int \phi_k(\mathcal{H} - W)\phi_l \, d\tau = (\mathcal{H}^{\circ}_{kl} - S_{kl}W + X_{kl}H^2) \exp(2\pi e i H D_{kl}/hc) \tag{3}$$

Here, $\mathcal{H}$ is the complete one-electron Hamiltonian operator which applies in the presence of the magnetic field, and $\mathcal{H}^{\circ}$ is the corresponding operator which would apply in the absence of such a field. In addition, D_{kl} is the area of the triangle defined by the origin and by the points $\mathbf{r}_k$ and $\mathbf{r}_l$; it is to be taken with positive or negative sign if these three points lie in, respectively, a clockwise or a counterclockwise order. Clearly, D_{kk} is identically equal to zero for all k, and $D_{kl} = -D_{lk}$. Finally, S_{kl} and X_{kl} are defined by the identities

$$S_{kl} \equiv \int \phi_k \phi_l \, d\tau \tag{4}$$

$$X_{kl}H^2 \equiv (e^2/2mc^2) \int \phi_k(\mathbf{A} - \mathbf{A}_k)^2 \phi_l \, d\tau \tag{5}$$

where m is the mass of an electron.

If we now make the following substitutions and approximations

$$\left.\begin{aligned} \mathcal{H}^{\circ}_{kl} &= \alpha \text{ for } k = l \\ &= \gamma \text{ for } \phi_k \text{ and } \phi_l \text{ adjacent} \\ &= 0 \text{ otherwise} \end{aligned}\right\} \tag{6}$$

$$\left.\begin{aligned} S_{kl} &= 1 \text{ for } k = l \\ &= S \text{ for } \phi_k \text{ and } \phi_l \text{ adjacent} \\ &= 0 \text{ otherwise} \end{aligned}\right\}, \tag{7}$$

$$\left.\begin{aligned} X_{kl} &= -(1/2)\chi_0 \text{ for } k = l \\ &= -(1/2)T \text{ for } \phi_k \text{ and } \phi_l \text{ adjacent} \\ &= 0 \text{ otherwise} \end{aligned}\right\} \tag{8}$$

$$x \equiv \frac{\alpha - W - (1/2)\chi_0 H^2}{\gamma - SW - (1/2)TH^2} \tag{9}$$

and

$$p_{kl} \equiv (2\pi e H D_{kl})/hc \tag{10}$$

the secular determinant can be put into a simple and convenient form. Each diagonal element is then equal to x; each non-diagonal element connecting adjacent orbitals ϕ_k and ϕ_l is equal to either $\exp(ip_{kl})$ or to $\exp(-ip_{kl})$; and all other elements are equal to zero. The Hermitian character is maintained, since all diagonal elements are real, and since $p_{kl} = -p_{lk}$.

Solution of the secular equation gives x as a function of H, expressible as a power series

$$x = x_0 + x_1 H + x_2 H^2 + \ldots. \tag{11}$$

The coefficients x_0, x_1, x_2, . . are independent of H, but they do in general depend upon the areas D_{kl}.

From equation (9) we find that

$$W = \frac{\alpha - x\gamma - (1/2)\chi_0 H^2 + (1/2)TH^2 x}{1 - Sx}. \tag{12}$$

In the original treatment of London,[1] both integrals S and T were assumed to be equal to zero. With this assumption, equation (12) becomes

$$W = \alpha - (1/2)\chi_0 H^2 - x\gamma = \alpha - (1/2)\chi_0 H^2 - (x_0 + x_1 H + x_2 H^2 + \ldots)\gamma \tag{13}$$

so that the magnetic susceptibility χ, due to an electron in the molecular orbital which has the energy W, is

$$\chi = -\left.\frac{\partial^2 W}{\partial H^2}\right|_{H=0} = \chi_0 + 2x_2\gamma. \tag{14}$$

If the molecular orbital ψ happens to be identical with one of the atomic orbitals ϕ_k, both x and x_2 are equal to zero. Hence χ_0 is merely the susceptibility of an electron in an unmodified atomic p orbital. On the other hand, if the molecular orbital is of the two-center type involved in a localized bond between two definite atoms, x is found always to be independent of H, so that x_2 must be equal to zero. Here again, therefore, the susceptibility is simply χ_0. Consequently, it seems reasonable to suppose that, with a planar molecule like that of benzene, the sum of the terms $2x_2\gamma$ for the occupied molecular orbitals, which are responsible for the unlocalized π bonds, is equal to the diamagnetic anisotropy $\chi_\perp - \chi_\parallel$ or, in other words, to the difference between the diamagnetic susceptibilities in the directions perpendicular to and parallel to the plane of the molecule. London showed that, when the diamagnetic anisotropies of several different planar aromatic hydrocarbons are calculated in this way, satisfactory agreement with experiment can be obtained if the resonance integral γ is assigned a value of about -4.4 e.-v. It is to be noted that γ is therefore not only negative, as it is generally assumed to be in the calculations of other properties such as the resonance energy, but also of reasonable magnitude. Moreover, γ appears to be at least approximately constant from molecule to molecule. The results of London's calculations are listed in the second column of table 1 on page 370.

Somewhat after London described his original treatment of diamagnetic anisotropy, Brooks[2] considered the modifications that are required if the overlap integral S is assigned a value different from zero. The following discussion, although superficially rather different from that of Brooks, seems to be at least approximately equivalent to his and, in addition, it brings out more clearly the nature of the approximations which are made. Let us introduce a new parameter y, which is related to the x of equation (9) by the identity

$$y \equiv \frac{x}{1 - Sx} \tag{15}$$

so that

$$x \equiv \frac{y}{1 + Sy}. \tag{16}$$

Substitution of 16 into 12 then gives

$$W = \alpha - (1/2)\chi_0 H^2 - y[\gamma - S\alpha + (1/2)S\chi_0 H^2 - (1/2)TH^2]. \tag{17}$$

Now, if we define

$$\beta \equiv \gamma - S\alpha \tag{18}$$

and if we assume that

$$T = S\chi_0 \tag{19}$$

we find that

$$W = \alpha - (1/2)\chi_0 H^2 - y\beta. \tag{20}$$

This resulting equation has the same form as 13, and differs from it only in the interpretation of the term $y\beta$, which here replaces the term $x\gamma$ of the earlier equation. From this point, we can therefore proceed as before by setting

$$y = y_0 + y_1 H + y_2 H^2 + \ldots \tag{21}$$

(cf. equation (11)) and by then obtaining the values of y_0, y_1, y_2, ... from equations 11 and 15. The diamagnetic anisotropy then of course becomes the sum of the terms $2y_2\beta$ for the occupied orbitals. The agreement between the observed anisotropies and the values calculated in this way is essentially as good as with London's earlier treatment.[3] The anisotropies calculated by Brooks with $S = 1/4$ are listed in the third column of table 1 on page 370.

Since the above-described London-Brooks treatment of diamagnetic anisotropy seems to give satisfactory results with simple aromatic hydrocarbons, an attempt was made to extend it to some rather different systems. From a study of the apparent magnetic susceptibilities of some highly dissociated hexaarylethanes $Ar_3C—CAr_3$, Selwood and Dobres[4] concluded that the magnitudes of the diamagnetic contributions to the total susceptibilities of the triarylmethyl radicals are probably much larger than Pascal's relations would lead one to expect. This conclusion seems entirely reasonable since the increase in the extent of the conjugated system in the transition from an ethane of this type to the corresponding pair of radicals should result in an increase in the diamagnetic anisotropy, so that, for example, the anisotropy of triphenylmethyl should be greater than three times that of benzene. When the calculation (by the original method of London) was made, however, the result was quite different from the expectation, since the calculated anisotropy of triphenylmethyl was only about 2.6 times that of benzene. (Cf. table 1, page 370.) This discrepancy might be due to the fact that the calculation was based on the presumably incorrect assumption that the radical is planar. Such an explanation, however, seems unlikely for, since any departure from planarity must decrease the effectiveness of the conjugation of the rings with the central carbon atom and with one another, the calculation for a non-planar radical would be expected to give a result which is intermediate between that for the planar radical and that for isolated benzene molecules.

Since the treatment of triphenylmethyl was therefore far from satisfactory, we considered it desirable to treat some other, less complex systems in which, as in triphenylmethyl, the conjugation is dependent on the

presence of an unsaturated atom that is not part of any ring. A simple example of such a system is provided by the carbonate ion, which has the required planar configuration, and which exhibits a small but appreciable diamagnetic anisotropy of the same type as does benzene. The presence of oxygen, instead of carbon, atoms in this molecule and the existence of the ionic charge doubtless require the use here of different values for the parameters S, α, β, γ, and χ_0, but it was hoped that qualitative information could nevertheless be obtained. The calculations, however, showed that, no matter what values were assigned to these parameters, the calculated anisotropy was exactly zero. Some modification of the treatment was therefore demanded.

Our first idea was that the calculation ought to lead to the desired anisotropy if we "completed the electric circuits" by letting the various integrals connecting different atomic orbitals (e.g., S, γ, T) have non-vanishing values even when the two orbitals in question were not adjacent. Although several additional parameters would therefore be required, we again hoped that we might be able to obtain qualitatively satisfactory information. The calculations, however, led to the completely unexpected result that the anisotropy, although now no longer zero, was in the wrong direction. In fact, with any reasonable choice of values for the parameters, the ion was found (contrary to experiment) to be less diamagnetic in the direction perpendicular to its plane than in one parallel to that plane Although this error might be due to the fact that the carbonate ion differs from a hydrocarbon in having a larger number of π electrons than of π orbitals, such an explanation seems unlikely since a similar situation was encountered also with butadiene. Although no experimental data are available, one would naturally suppose that this latter substance must have its greatest diamagnetic susceptibility perpendicular to the plane of the molecule. As with the carbonate ion, the calculation leads to zero anisotropy if only the interactions between adjacent orbitals are considered, and to an anisotropy in the wrong or, at any rate, in the unexpected direction if more distant interactions are also included. In view of the serious discrepancies to which it thus led, this method for generalizing the London-Brooks treatment was now abandoned, and we returned to the simpler procedure of neglecting all interactions between non-adjacent orbitals.

Our next, and final, attempt to improve the treatment of diamagnetic anisotropy was slightly more successful than the first, but it still left much to be desired. As was noted above, the calculations in which the overlap integral S was not neglected were based on the assumption expressed in equation (19). Since all the pertinent integrals can, however, be readily evaluated, it now seemed worth while to see how nearly correct this assumption really is. If the atomic orbitals ϕ_k have the form[5]

$$\phi_k = Nr \sin\theta \cos\phi \exp(-1.625r/a_H) \quad (22)$$

where N is a normalization constant and a_H is the radius of the first Bohr orbit, the overlap integral S for two adjacent orbitals at a distance of 1.4 A is found[5] to be about one-fourth. On the other hand, the ratio T/χ_0 is found then to be more nearly equal to three-fourths. Thus, we see that, in assuming T to be equal to zero, London neglected a term nearly as important as the one, χ_0, which he retained; and, moreover, that the assumption (19) is rather inaccurate. We therefore have reason to hope that the calculations could be improved by use of a better value for the ratio T/χ_0.

By combining equations (17), (18), and (21), we obtain

$$W = \alpha - (1/2)\chi_0 H^2 - (y_0 + y_1 H + y_2 H^2 + \ldots) \times [\beta + (1/2)(S\chi_0 - T)H^2] \quad (23)$$

so that, if we set $S = 1/4$ and $T/\chi_0 = 3/4$,

$$W = \alpha - (1/2)\chi_0 H^2 - (y_0 + y_1 H + y_2 H^2 + \ldots)(\beta - \chi_0 H^2/4) \quad (24)$$

Consequently,

$$\chi_\perp = -\left.\frac{\partial^2 W}{\partial H^2}\right|_{H=0} = \chi_0(1 - y_0/2) + 2y_2\beta \quad (25)$$

The diamagnetic anisotropy, which is now

$$\chi_\perp - \chi_\parallel = -(1/2)\chi_0 y_0 + 2y_2\beta \quad (26)$$

therefore differs by the term $-(1/2)\chi_0 y_0$ from the one based on equation (20). Since χ_0 and y_0 are both negative, the effect of this term is to increase the magnitude of the anisotropy that is calculated for the entire molecule by summation of the contributions $\chi_\perp - \chi_\parallel$ for the individual occupied orbitals. Hence, the correction is at any rate in the right direction.

In order to apply equation (26) to a specified molecule, we need not only to determine the values of the constants y_0 and y_2 by solution of the secular equation, but also to estimate the values of the parameters χ_0 and β. The first of these latter quantities can be directly evaluated by integration of the defining equation (5) (cf. also equation (8)); thus, if ϕ_k is assumed to have the same form as before (cf. equation (22)), χ_0 is found to be equal to -1.35×10^{-6}. The remaining parameter, β, is then chosen so as to give agreement between the calculated and observed anisotropies of benzene. In this way, β is found to be equal to -4.20 e.-v.

The results of several calculations based on equation (26) are given in the fourth column of table 1. For the aromatic hydrocarbons, the present treatment is approximately as satisfactory as, but not significantly better than, those of London and of Brooks. For the carbonate ion, the calculated anisotropy is at least in the right direction but, with any reason-

able choice of the additional parameters that are introduced by the difference between oxygen and carbon, it is much too small. For butadiene and for the allyl radical, the calculated anisotropies are in the expected direction, but no experimental values are available for comparison. Finally, for triphenylmethyl, the calculated anisotropy is slightly greater

TABLE 1

CALCULATED AND OBSERVED DIAMAGNETIC ANISOTROPIES

	MAGNITUDE OF ANISOTROPY ($\times 10^6$)			
SUBSTANCE	LONDON[a]	BROOKS[b]	PRESENT TREATMENT[c]	OBSERVED[d]
Benzene	(54)	(54)	(54)	54
Naphthalene	118	120	117	114
Anthracene	186	191	186	183
Phenanthrene	172	165	163	166
Pyrene	241	257	242	233
Biphenyl	103[d]	103[e]	104	119
Stilbene (trans)	97[f]	...	103	124–160[g]
Triphenylmethyl	139[f, h]	...	141[h]	>162[i]
Ethylene	0[f]	0[f]	1.1	...
Allyl radical	0[f]	0[f]	1.4	...
Butadiene	0[f, h]	0[f]	2.3	...
Cyclopentadiene	0[f, h]	0[f]	2.3	...
Carbonate ion	0[f]	0[f]	1.6[j]	4[k]

[a] Except as noted, these values are taken from London, F., *J. phys. radium*, **8**, 397 (1937).

[b] Except as noted, these values are taken from Brooks, H., *J. Chem. Phys.*, **9**, 463 (1941).

[c] Calculated with $S = {}^1/_2$, $T = 3\chi_0/4$, $\chi_0 = -1.35 \times 10^{-6}$, $\beta = -4.20$ e.-v.

[d] Corrected value, obtained by the present authors. London's original value was 119. We have been unable to confirm the value of 101 given by McWeeny (cf. footnote 3).

[e] Corrected value, obtained by the present authors. Brooks' original value was 185. We have been unable to confirm either the value of 106 given by McWeeny in his first paper or the different value of 101 given by the same author in his second paper (cf. footnote 3).

[f] Calculated by the present authors.

[g] Lonsdale, K., *Proc. Roy. Soc.* (*London*), **A171**, 541 (1939).

[h] Calculated for a planar molecule.

[i] Cf. Selwood, P. W., and Dobres, R. M., *J. Am. Chem. Soc.*, **72**, 3860 (1950).

[j] Calculated on the assumption that α for an oxygen atom is the same as for a carbon atom. Agreement with experiment could be obtained only by making S negative or T greater than χ_0. It is not inconceivable that better results could have been obtained by using different values for the two different integrals α.

[k] Krishnan, K. S., *Phys. Rev.*, **38**, 833 (1931).

than that found by the London treatment but, since it is less than three times that of benzene, it is still much too small.

Summary and Conclusion.—In the present discussion of the molecular-orbital calculation of diamagnetic anisotropy (with neglect of all inter-

actions between non-adjacent orbitals), three different sets of assumptions regarding the values of the integrals S (equation (7)) and T (equations (5) and (8)) have been considered. (1) The original treatment, due to London, results when both S and T are arbitrarily set equal to zero. (2) A presumably more accurate treatment, which is at least approximately equivalent to the one due to Brooks, results when S is given the value $^1/_4$ that is obtained by integration; and when T is assumed to be equal to $S\chi_0$. (3) A presumably still more accurate treatment results when S, T and χ_0 are assigned the values ($^1/_4$, $3\chi_0/4$ and -1.35×10^{-6}, respectively) that are obtained by integration.

For aromatic hydrocarbons, all three treatments lead to calculated anisotropies in fairly good agreement with one another and with experiment. For the carbonate ion and for the triphenylmethyl radical, no one of the methods is at all satisfactory, since all give much too small anisotropies. For the allyl radical and for butadiene, the last method gives anisotropies in the expected directions, but a comparison with experiment is prevented by lack of data. Of the three methods, no one is therefore generally satisfactory, but the last is perhaps slightly better than the other two.

If, in the calculations of the London type, interactions between non-adjacent orbitals are not neglected, the calculated anisotropies are in the wrong direction for the carbonate ion and probably also for butadiene.

* This paper was prepared in connection with a conference on "Quantum-Mechanical Methods in Valence Theory" held at Shelter Island, September 7–10, 1951, under the auspices of the NATIONAL ACADEMY OF SCIENCES (see Summary of the Conference in a subsequent issue of these PROCEEDINGS).

[1] London, F., *J. phys. radium*, **8**, 397 (1937).

[2] Brooks, H., *J. Chem. Phys.*, **9**, 463 (1941).

[3] For one compound, biphenyl, Brooks reported[2] that his method gives a calculated anisotropy which is about 55% too high. This discrepancy, apparently arises from the fact that, in his calculations, Brooks used the solutions of the secular equation which are listed by London on page 408 of his original paper [1] As was first noted by R. McWeeny [*J. Chem. Phys.*, **17**, 1341 (1949)], however, there are numerical errors in these latter figures. When the necessary corrections are made, the calculations by the methods of London and of Brooks are brought into essential agreement with each other, although each then leads to an anisotropy which is rather too low. (Cf. table 1.)

[4] Selwood, P. W., and Dobres, R. M., *J. Am. Chem. Soc.*, **72** 3860 (1950). However, cf. Chu, T., and Weissman, S. I., *Ibid.*, **73**, 4462 (1951).

[5] Mulliken, R. S., Rieke, C. A., Orloff, D., and Orloff, H., *J. Chem. Phys.*, **17**, 1248 (1949).

π-ELECTRONS AND σ-ELECTRONS*

BY C. A. COULSON,† N. H. MARCH‡ AND S. ALTMANN§

WHEATSTONE PHYSICS DEPARTMENT, KING'S COLLEGE, LONDON, ENGLAND

Communicated by Robert S. Mulliken, Feb. 26, 1952

In order to elucidate the distinction between π- and σ-electrons we have calculated[1] the density of the charge cloud in benzene. Two methods have been used for this purpose:

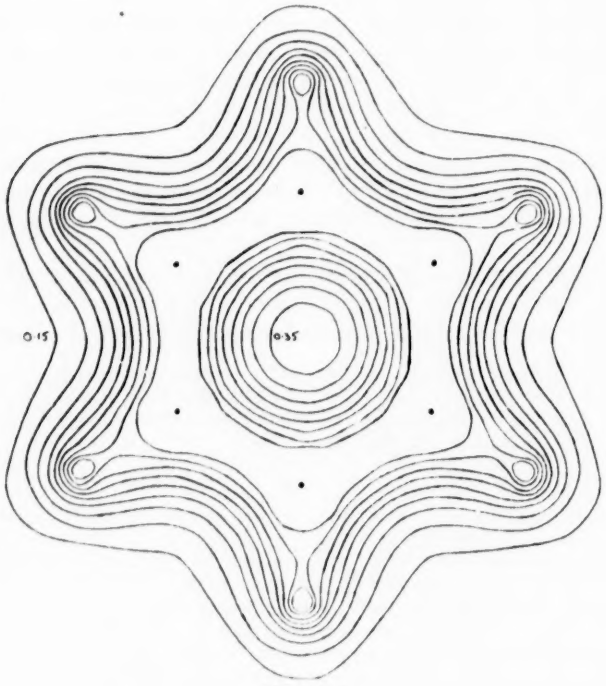

FIGURE 1

Calculated electron density in benzene in the plane of the ring; m. o. method. Dots denote carbon atoms. Contours at intervals of 0.20 e/A^3.

(*a*) Conventional molecular-orbital method, without configuration interaction.

(*b*) Thomas-Fermi statistical method.

The first of these methods distinguishes between σ- and π-electrons:

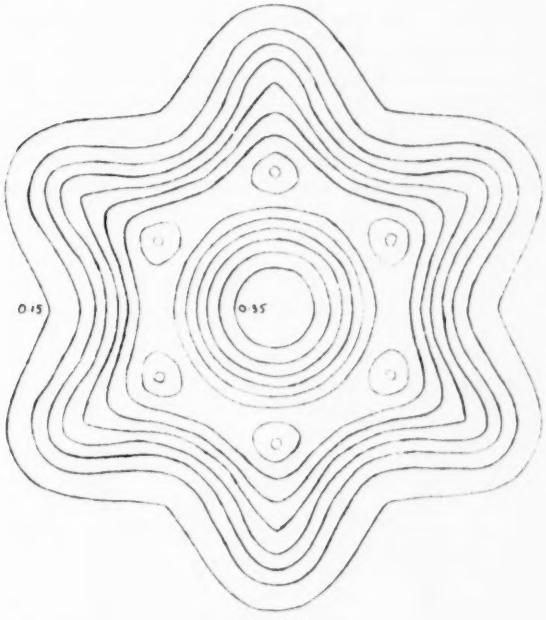

FIGURE 2

Same as figure 1, except that the plane is drawn parallel to the plane of the nuclei a distance 0.35 A. above it.

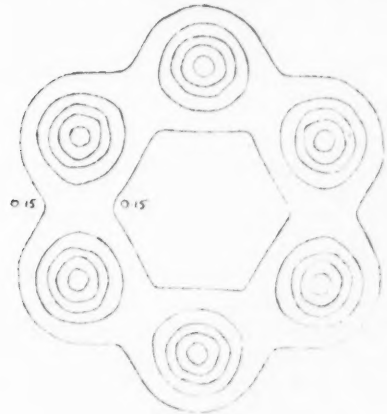

FIGURE 3

Same as figure 2, except that only the π-electrons are considered.

the second does not. The wave functions of the first (including the hydrogen atoms) are simple to write down. The solution of the T. F. equation required in the second is a matter of some difficulty, and was obtained by an extension of the technique used by Hund[2] in some early work on N_2. It seems reasonable to suppose that, although there are obvious approximations and assumptions, the two methods should give a good general description of the charge density except, perhaps, at large distances from the centers of the atoms.

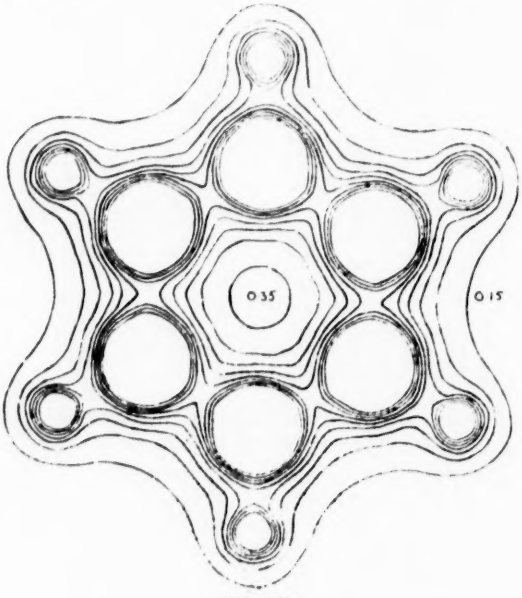

FIGURE 4

Calculated electron density in benzene in the plane of the ring. Thomas-Fermi method. Contours at intervals of 0.20 e/A^3. This may be compared with figure 1.

Numerical results are shown in figures 1–5, in which the contours of charge density are plotted both in the plane of the molecule, and in a parallel plane. The height of this plane above the plane of the ring (0.35 A.) was chosen so as to emphasize as much as possible the contribution from the π-electrons. This contribution is obviously zero in the molecular plane.

A study of figures 1–5 reveals the following:

(1) There is general agreement, from figures 1 and 4, that the charge

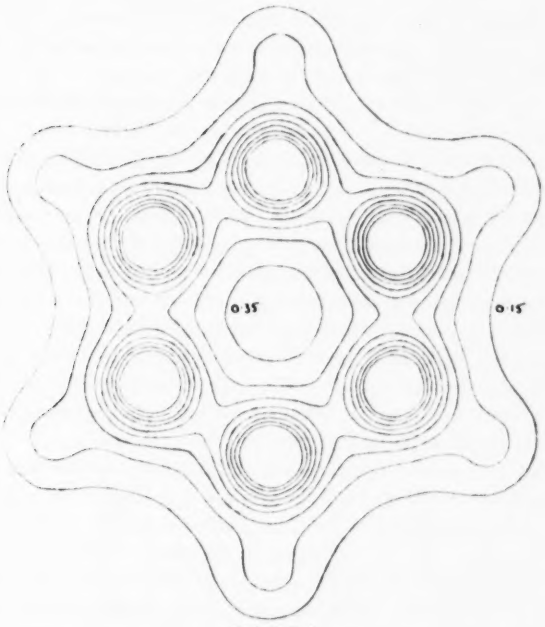

FIGURE 5

Same as figure 4, except that the plane is drawn parallel to the plane of the nuclei a distance 0.35 A. above it. This may be compared with figure 2.

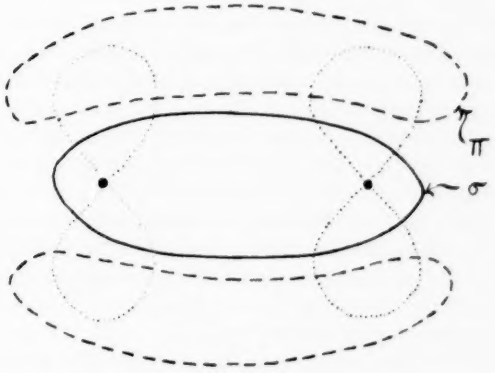

FIGURE 6

Conventional σ and π clouds.

cloud has a considerable density in the center of the ring, and that this density falls to about one-fifth of its value at the mid-point of the C—C bonds. This is in general agreement with the accurate x-ray analysis of Robertson[3] for naphthalene. (It would be highly desirable to have a good x-ray analysis of benzene.)

(2) The value of the cloud density along the C—C "bonds" suggests that these are slightly "bent" in toward the center of the hexagon.

(3) Most significant of all, however, is the situation in the parallel plane. For as figures 2 and 3 illustrate, there is only a very tiny region, immediately above the six-carbon nuclei, where the contribution from the π-electrons is as great as that of the σ-electrons. This means that the π-electrons may be thought of as practically "submerged" in a larger sea of σ-electrons. Such a situation is disguised from us by our conventional drawings (Fig. 6) of σ- and π-bonds. It is clear from the present analysis that, so far as their spatial character is concerned, the double-streamers of the π-cloud do not lie outside, or separate from, the σ-cloud.

This conclusion regarding the density makes one wonder how it is that so much progress has been made in organic chemistry by considering the π-electrons by themselves. It will be recalled that discussions of chemical reactivity, particularly for substitution reactions, and of polarizability, bond length, resonance moment, ultra-violet absorption, etc., are almost always given in terms of π-electrons only. It can only be concluded that such calculations are less surely based, and therefore more empirical, than we had previously thought. The σ- and the π-electrons cannot be clearly separated in this way.

This conclusion leads us to consider in what way we can incorporate such non-separability in our calculations. With this end in view one of us[4] has discussed ethylene, using an essentially valence-bond (HLPS) attack. More precisely we have considered the difference in ground and excited states between the predictions obtained (*a*) by treating the molecule as a purely π-electron problem—the so-called Hückel approximation—and (*b*) by introducing the σ-electrons and allowing the pairing of a σ- and a π-electron—the so-called σ-π resonance effect. In this work the full 12-electron Hamiltonian was used, and all the integrals except one purely atomic one, were either evaluated accurately, or, in three or four cases, values had to be estimated. The 1s, or K-shell, electrons were compressed into the nuclei. Although this will certainly affect the total energy, it is most unlikely to affect resonance and excitation energies to any significant extent. In this work ionic structures were neglected. A summary of the conclusions obtained is as follows:

(1) No *ab initio* calculation of the π-π exchange integral, such as those of Griffing[5] and Van Dranen,[6] is able to give a satisfactory value unless the full π-σ Hamiltonian is used. In our case, the π-π exchange inte-

gral is -2.7 e. v. obtained without the use of any empirical information except the C═C distance and the one atomic term referred to before. The σ-σ exchange integral is numerically rather smaller than the π-π exchange integral. The π-σ exchange integral is $+0.97$ e. v., with opposite sign to the π-π and σ-σ exchange integrals.

(2) A good deal of the stability of ethylene resides in π-σ resonance. Something like $^1/_2$ e. v. results in this way. It can be shown, however, that this resonance is not likely to be significant for the calculation of resonance energies. For these latter are differences between two energy values, in both of which the π-σ resonance is expected to be much the same.

(3) The situation in excited states, however, is quite different. In the simple HLPS scheme, many difficulties had already become apparent, and they had chiefly been assigned to lack of consideration of ionic structures. But it now appears much more probable that they lie in the failure of the Hückel approximation. A particularly striking illustration arises in ethylene. If we consider only the π-electrons, it is a 2-electron problem, and there is no excited singlet level arising from $p\pi$ atomic orbitals in a covalent wave function. Introduction of the σ-electrons of the C—C bond (without, at this stage, introducing the C—H σ-electrons) makes it a 4-electron problem which admits, in the same conditions, one excited singlet.

In our calculations we have introduced all 12 valence electrons, and incorporated canonical structures involving not more than two π-σ pairings of the electrons. Our suggested description of the spectrum in the ultraviolet now becomes as follows:

First there is a forbidden $^1A_{1g} - {}^1A_{1g}$ band, to be identified with the observed faint absorption starting at about 2000 A. Then there is a band, of allowed transition, but double, given by the symmetries $^1A_{1g} \rightarrow {}^1B_{1u}$ and $^1A_{1g} \rightarrow {}^1B_{2u}$, separated by about a calculated 0.2 e. v. (experimental value 0.06 e. v.). These bands are identified with the strong absorption at 1600 A. The double nature of this band has been recognized experimentally for some time, but it does not yet appear to have received any completely satisfactory interpretation. (The only current interpretation[7, 8] of which we are aware interprets these two bands as 0,0 and 0,2 transitions, with no sign of 0,4. This interpretation deals nicely with the isotope effect, but is in difficulty with the intensities.) The calculated energy difference between the first (forbidden) band and the second (allowed) band is 1.4 e. v. (experimental 1.1 e. v.). The chief objection to this new treatment is that the energies of the bands are about 1.5 e. v. higher than those given by the ASMO method.

All this suggests rather strongly that our rather heavy concentration on π-electrons to the almost complete exclusion of σ-electrons is by no means satisfactory, more particularly in the excited states.

It may be added in conclusion that Altmann has discussed alkyl-ethyl-

enes in the same way. A red shift in the tetra methyl compound is obtained, which is estimated at 0.9 e. v., to be compared with the experimental 1.1 e. v. This is particularly interesting in view of the fact that no explicit mention of hyperconjugation occurs in this treatment. Conventionally hyperconjugation is usually regarded as the source of this red shift. It is possible that Altmann's treatment does implicitly include a degree of σ-hyperconjugation. It does not, however, appear to include any π-hyperconjugation.

This method could be applied to benzene, but it would be very clumsy to apply to larger systems.

We should like to acknowledge the benefit of correspondence and discussion with Professor R. S. Mulliken.

* This paper was prepared in connection with the recent Academy conference on "Quantum-Mechanical Methods in Valence Theory," of which a Summary will be found in a subsequent issue of the PROCEEDINGS.

† Now at Mathematical Institute, Oxford.

‡ Now at Physics Department, The University, Sheffield.

§ Now at Chemistry Department, The University, Buenos Aires.

[1] March, N. H., *Acta Crystall.* (in press).

[2] Hund, F., *Z. Phys.*, **77,** 12 (1932).

[3] Abrahams, S. C., Robertson, J. M., and White, J. G., *Acta Crystall.*, **2,** 238 (1949).

[4] Altmann, S. L., *Proc. Roy. Soc.*, (*London*), **A210,** 327, 343 (1952).

[5] Griffing, V., *J. Chem. Phys.*, **15,** 421 (1947).

[6] Van Dranen, J., and Ketelaar, J. A. A., *Ibid.*, **18,** 1125 (1950).

[7] Mulliken, R. S., *Rev. Mod. Phys.*, **14,** 265 (1942).

[8] Mulliken, R. S., and Roothaan, C. J. A., *Chem. Rev.*, **41,** 219 (1947).

ON THE ANOMALOUS TRANSPARENCY OF THICK CRYSTALS TO X-RAYS

BY W. H. ZACHARIASEN

DEPARTMENT OF PHYSICS, UNIVERSITY OF CHICAGO, AND ARGONNE NATIONAL LABORATORY

Communicated February 8, 1952

The anomalous transparency of thick, perfect crystals to a beam of x-rays producing "Laue case" diffraction was first observed by G. Borrmann.[1] The complete solution for Laue case diffraction in a perfect crystal was first reported in 1945,[2] and the theoretical explanation of Borrmann's observation is contained in the equations giving this solution.[3] The writer was not aware of Borrmann's experiment, and the significance of the equations as relating to anomalous transparency was not pointed out.

The first quantitative experimental studies of the effect have recently

been carried out by Borrmann[4] and by H. N. Campbell,[5] both workers using calcite crystals and copper K radiation, but different experimental procedures. Neither Borrmann nor Campbell were familiar with the earlier theoretical work of the writer, and hence they did not compare their experimental results with theoretical predictions. It is the purpose of the present paper to examine to what extent there is quantitative agreement between theory and experiment.

The Ideal Single Crystal Diffraction Pattern.—In order to avoid unnecessary duplication the notation of reference 2 will be used without further explanation, and numbered equations in reference 2 will be referred to by means of a prefix Z. The pertinent equations are inavoidably rather complicated; but some simplification is attained if the considerations are restricted to the symmetrical Laue case (i.e., $b = +1$) and to strong reflections in centro-symmetrical crystals. Under these circumstances one has (see equations Z-3.123 and Z-3.177)

$$\sqrt{b\psi_H\psi_{\bar{H}}} \approx |\psi_H'|(1 + i\kappa) \tag{1}$$

where $|\kappa| \ll 1$. The intensity distribution for the diffracted wave is then given by equation Z-3.184. The corresponding explicit formula for the transmitted wave is obtained from equation Z-3.131 using equation 1 above. The result is

$$\frac{I_e^0}{I_0^e} = \frac{e^{-\mu_0(t_0/\gamma_0)}}{2(1 + y^2)}\left\{(1 + 2y^2)\cosh\left(\frac{2\kappa A}{\sqrt{1 + y^2}}\right) + 2y\sqrt{1 + y^2}\sinh\left(\frac{2\kappa A}{\sqrt{1 + y^2}}\right) + \cos\left(2A\sqrt{1 + y^2}\right)\right\}. \tag{2}$$

Inspection of equations 2 and Z-3.184 shows that anomalous transparency does not become important unless $|\kappa| A > 1$. This condition is equivalent to the following

$$\mu_0 t_0 \geq \mu_H t_0 > 1 \tag{3}$$

where μ_H is the Fourier coefficient given in equation Z-3.99 and t_0 the thickness of the crystal plate.

When $|\kappa| A > 3$ equation Z-3.188 can be used instead of equation Z-3.184, and equation 2 can be rewritten in the form

$$\frac{I_e^0}{I_0^e} \approx \frac{e^{-\mu_0(1 - e/\sqrt{1 + y^2})t_0/\gamma_0}}{4(1 + y^2)}\{1 + 2y^2 + 2y\sqrt{1 + y^2}\}. \tag{4}$$

Indeed, still further simplification is justifiable. The half width at half maximum, W_y, is seen to be

$$W_y \approx \sqrt{\frac{2 \log 2}{\epsilon \mu_0 \dfrac{t_0}{\gamma_0}}} \ll 1 \tag{5}$$

for both transmitted and diffracted waves. The intensity ratio is thus significantly different from zero (or $e^{-\mu_0(t_0/\gamma_0)}$) only in the range $|y| \ll 1$. Equations Z-3.188 and 4 may accordingly be approximated as follows

$$\frac{I_H}{I_0^e} \approx \frac{I_e^0}{I_0^e} \approx \frac{1}{4} e^{-\mu_0(1 - \epsilon/\sqrt{1 + y^2})t_0/\gamma_0}. \tag{6}$$

Clearly equation 6 becomes unreliable when $|y|$ far exceeds the half width value.

According to equation 6 the intensity associated with the transmitted wave is practically equal to that associated with the diffracted wave. Both Borrmann and Campbell found this to be true for sufficiently perfect crystal specimens.

The formula of equation 6 gives the ideal single crystal diffraction pattern. However, this pattern cannot be observed since no point source of monochromatic x-rays is experimentally attainable. As a consequence Borrmann's and Campbell's results cannot be directly compared with equation 6.

Comparison with Borrmann's Results.—In Borrmann's work the condition for symmetrical Laue case is not exactly fulfilled. For $\lambda = 1.54$ Å one calculates $b = 1.15$. However, it is justifiable to assume that equation 6 can still be used without introducing appreciable errors.

The divergent radiation from the copper target of an x-ray tube served as the incident beam. The transmitted and diffracted beams were recorded on a photographic plate, and intensity ratios were measured by means of a photometer. The x-rays reaching the center of the transmitted or diffracted line on the photographic plate do not come only from the center of the focal spot (corresponding to $y = 0$), but also from neighboring points of the focal spot (corresponding to $y \gtrless 0$). Accordingly the maximum intensity ratio which is measured by Borrmann represents the integrals

$$R_H = \int \frac{I_H}{I_0^e} dy = \int \frac{I_e^0}{I_0^e} dy = R_e^0. \tag{7}$$

With very good approximation the integration yields

$$R_H = R_e^0 \approx \left(\frac{8\epsilon\mu_0 t_0}{\pi\gamma_0}\right)^{-1/2} e^{-\mu_0(1 - \epsilon)t_0/\gamma_0}. \tag{8}$$

The effective absorption coefficient μ_I measured by Borrmann should thus be given by the equation

$$\mu_I = \mu_0 \left\{ 1 + \frac{\log\left(\frac{8\epsilon\mu_0 l_0}{\pi\gamma_0}\right)}{2\mu_0 \frac{l_0}{\gamma_0}} - \epsilon \right\}. \qquad (9)$$

The value of μ_0 for $\lambda = 1.54$ Å for calcite is 193 cm.$^{-1}$. For the cleavage plane $\gamma_0 = \cos\theta = 0.967$. Using the numerical values of equations Z-3.198 and Z-3.200 one finds $\epsilon = 0.896$ (0.958 for normal and 0.834 for parallel polarization). Borrmann's observed values of μ_I and those calculated from equation 9 are given in table 1.

TABLE 1

CALCULATED AND OBSERVED VALUES OF μ_I

l_0 in cm.	Calculated, cm.$^{-1}$	Observed, cm.$^{-1}$
0.212	30.5	30.2
0.271	28.6	27.7

Comparison with Campbell's Results.—In Campbell's experiments a double crystal diffraction procedure was used. The relative orientation of the two crystals corresponded to the so-called (−) position (see p. 148 of reference 2). Diffraction in the first calcite crystal corresponded to the symmetrical Bragg position and in the second crystal to the symmetrical Laue case. By means of Geiger counters Campbell measured the ratio of the power transmitted or diffracted by the second crystal to that incident upon the second crystal as function of the setting of the second crystal. The rocking curve which Campbell obtained in this manner accordingly corresponds to the function P given in equations Z-3.210 and Z-3.216. The power ratio measured exactly at the Bragg angle is the quantity $P(0)$.

The half width value for the diffraction pattern from the first crystal is large compared to that for the second crystal ($W_y \approx {}^3/_4\sqrt{2}$ for the first crystal while $W_y \ll 1$ for the second crystal according to equation 5). Accordingly equation Z-3.216 for $y' = 0$ can be simplified to

$$P(0) = CR_H = CR_e^0 \qquad (10)$$

where $C = \frac{1}{R_H}\frac{I_H}{I_0}(0)$ for the first crystal. The quantity C will have very nearly the same value as for the Darwin solution so that $C = {}^3/_8 \approx e^{-1}$. The power ratio at the center of the rocking curve obtained by Campbell is thus given by

$$P(0) = e^{-\mu_0} \left\{ 1 + \frac{\log\left(\frac{8\epsilon\mu_0 l_0}{\pi\gamma_0}\right) + 2}{2\mu_0 \frac{l_0}{\gamma_0}} - \epsilon \right\}^{\frac{l_0}{\gamma_0}} \qquad (11)$$

and the effective absorption coefficient by

$$\mu = \mu_0 \left\{ 1 + \frac{\log\left(\frac{8\epsilon\mu_0 t_0}{\pi\gamma_0}\right) + 2}{2\mu_0 \frac{t_0}{\gamma_0}} - \epsilon \right\} \qquad (12)$$

Only Campbell's etched calcite crystal of thickness 0.040 cm. was of sufficiently high degree of perfection to justify comparison between theory and experiment. The experimental data given by Campbell correspond to $\mu = 86$ cm.$^{-1}$, while the value calculated from equation 12 is $\mu = 79$ cm.$^{-1}$.

Excellent agreement between theory and the measurements of Borrmann and of Campbell is thus demonstrated.

[1] Borrmann, G., *Physik. Z.*, **42,** 157 (1941).

[2] Zachariasen, W. H., *Theory of X-Ray Diffraction in Crystals*, John Wiley & Sons, Inc., New York, 1945.

[3] The decreased absorption for the diffracted beam appears explicitly in equation 3.188 and in less obvious form for the transmitted beam in equation 3.131.

[4] Borrmann, G., *Z. Physik.*, **127,** 297 (1950).

[5] Campbell, H. N., *J. Appl. Phys.*, **22,** 1139 (1951).

ERRATA

In my article "On a Theorem of Hardy, Littlewood, Polya and Blackwell," the PROCEEDINGS, **37,** 826–831(1951), the formula in the middle of p. 828 should read

$$(E_a^+ + E_b^-) \cap (\mathrm{CON}^*) = \sum_{1 \leq j \leq n_b} (((E_a^+ \cap E_j^+) + [e_j]) \cap \bar{c}^{\perp}).$$

The third line below this should read

$$a = \sum_{1 \leq i \leq n_a} a_i + \sum_{1 \leq j \leq n_b} a'.$$

And the lemma on p. 829 should read

Lemma: For each non-zero $a - b \in (E_a^+ + E_b^-) \cap \mathrm{CON}^*$ there exists j, $1 \leq j \leq n_b$, and non-zero $d \in ((E_a^+ \cap E_j^+) + [e_j]) \cap \bar{c}^{\perp}$ such that $a - b \underset{1}{>} d \underset{1}{>} 0$

S. SHERMAN

INFORMATION TO CONTRIBUTORS

THE PROCEEDINGS is the official organ of the NATIONAL ACADEMY OF SCIENCES and of the NATIONAL RESEARCH COUNCIL for the publication of brief accounts of important current researches of members of the ACADEMY and of the COUNCIL and of other American investigators. The PROCEEDINGS will aim especially to secure prompt publication of orginal announcements of discoveries and wide circulation of the results of American Research among investigators in other countries and in all branches of science.

ARTICLES should be brief. The viewpoint should be comprehensive in giving the relation of the paper to previous publications of the author or of others and in exhibiting, where practicable, the significance of the work for other branches of science. Elaborate technical details of the work and long tables of data should be avoided, but authors should be precise in making clear the new results and should give some record of the methods and data upon which they are based.

MANUSCRIPTS should be prepared with a current number of the PROCEEDINGS as a model in matters of form, and should be typewritten in duplicate with double spacing, the author retaining one copy. Illustrations should be confined to text figures of simple character, though more elaborate illustration may be allowed in special instances to authors willing to pay for their preparation and insertion. Particular attention should be given to arranging tabular matter in a simple and concise manner.

REFERENCES to literature, numbered consectuvely, will be placed at the end of the article and short footnotes should be avoided. It is suggested that references to periodicals be furnished in some detail and in general in accordance with the standard adopted for the Subject Catalogue of the International Catalog of Scientific Literature, viz., name of author, with initials following (ordinarily omitting title of paper), abbreviated name of Journal, with place of publication, series (if any), volume, inclusive pages, year. For example: Montgomery, T. H., *J. Morph.*, Boston, 22, 731–815 (1911); or, Wheeler, W. M., *Konigsberg, Schr. physik. Ges.*, 25, 1–142 (1914).

MANUSCRIPTS will be accepted only from members of the Academy or from chairmen of the divisions of the National Research Council who will assume responsibility for the propriety and scientific standards of the paper and should be addressed to *Proceedings of the* NATIONAL ACADEMY OF SCIENCES, *2101 Constitution Ave., Washington, 25, D. C. Articles of which a member of the Academy is author or a co-author may extend to twelve printed pages, those by non-members only to six printed pages; space above these limits will be billed to the author at $17.50 per page.*

PROOF will ordinarily be sent,—with the understanding that charges for an author's correction shall be billed to him. Authors are therefore requested to make final revisions on the typewritten manuscripts.

REPRINTS should be ordered at the time of submission of manuscript. They will be furnished to authors at cost, approximately as follows:

SCHEDULE OF RATES FOR REPRINTS, CARRIAGE CHARGES EXTRA

No. Copies	50	100	150	200	250
4 pp.	4.68	5.20	5.85	6.50	7.15
8 pp.	9.10	10.40	11.70	13.00	14.30
Covers	3.45	4.42	5.41	6.39	7.37

CONTENTS

		Page
ASTRONOMY.—MAGELLANIC CLOUDS. III. DIFFERENTIAL ABSORPTION WITHIN THE LARGE CLOUD	*By Harlow Shapley and Virginia McKibben Nail*	281
ASTRONOMY.—MAGELLANIC CLOUDS. IV. ON THE PERIOD FREQUENCY ANOMALIES	*By Harlow Shapley and Virginia McKibben Nail*	286
GENETICS.—BREAKAGE OF CHROMOSOMES BY OXYGEN	*By Alan D. Conger and Lucile M. Fairchild*	289
GENETICS.—MUTANT ISOALLELES AT THE VERMILION LOCUS IN DROSOPHILA MELANOGASTER	*By M. M. Green*	300
MATHEMATICS.—THE PARACOMPACTNESS OF THE WEAK SIMPLICIAL COMPLEX	*By D. G. Bourgin*	305
MATHEMATICS.—SOME PROBLEMS INVOLVING PRIMITIVE ROOTS IN A FINITE FIELD	*By L. Carlitz*	314
MATHEMATICS.—ON ANALYTIC SURFACES WITH TWO INDEPENDENT MEROMORPHIC FUNCTIONS	*By Wei-Liang Chow and Kunihiko Kodaira*	319
MATHEMATICS.—COHOMOLOGY GROUPS OF ABELIAN GROUPS AND HOMOTOPY THEORY IV.	*By Samuel Eilenberg and Saunders MacLane*	325
MATHEMATICS.—THE APPARITION PROBLEM FOR EQUIANHARMONIC DIVISIBILITY SEQUENCES	*By L. K. Durst*	330
MATHEMATICS.—SOME THEOREMS ON PIECEWISE LINEAR EMBEDDING	*By V. K. A. M. Gugenheim*	333
MATHEMATICS.—PLANCHERELL FORMULA FOR THE 2 × 2 REAL UNIMODULAR GROUP	*By Harish-Chandra*	337
MATHEMATICS.—ON SOLUTIONS OF MAXWELL'S EQUATIONS IN AN EXTERIOR REGION	*By William K. Saunders*	342
MATHEMATICS.—ON COHOMOLOGY THEORIES	*By Chung-Tao Yang*	348
GEOPHYSICS.—ON THE EARLY CHEMICAL HISTORY OF THE EARTH AND THE ORIGIN OF LIFE	*By Harold C. Urey*	351
PHYSICS.—SOME COMMENTS ON THE LONDON-BROOKS TREATMENT OF DIAMAGNETIC ANISOTROPY	*By G. W. Wheland and Sheldon L. Mallow*	364
PHYSICS.—π-ELECTRONS AND σ-ELECTRONS	*By C. A. Coulson, N. H. March and S. Altmann*	372
PHYSICS.—ON THE ANOMALOUS TRANSPARENCY OF THICK CRYSTALS TO X-RAYS	*By W. H. Zachariasen*	378
ERRATA.—ON A THEOREM OF HARDY, LITTLEWOOD, POLYA AND BLACKWELL	*S. Sherman*	382